Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 145 (2010) 85–105
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Two-sided matching with interdependent values

Archishman Chakraborty [a], Alessandro Citanna [b], Michael Ostrovsky [c,*]

[a] *Schulich School of Business, York University, Toronto ON M3J 1P3, Canada*
[b] *HEC – Paris, 75381 Jouy-en-Josas, France*
[c] *Graduate School of Business, Stanford University, Stanford, CA 94305, USA*

Received 13 December 2007; final version received 17 March 2009; accepted 21 July 2009

Available online 2 September 2009

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. When colleges observe the entire matching, a stable matching mechanism does not generally exist. When colleges observe only their own matches, a stable mechanism exists if students have identical preferences over colleges, but may not exist if students have different preferences.

JEL classification: C78; D82

Keywords: Interdependent values; Stability; Matching

After the focal female had seen the model near one of the males, the model was removed and the focal female was allowed to choose between males. The focal female chose the male that had been beside the model female 17 out of 20 times.

Dugatkin [5], reporting the results of an experiment on female copying behavior during mate choice among Trinidadian guppies.

* Corresponding author.
E-mail addresses: achakraborty@schulich.yorku.ca (A. Chakraborty), citanna@hec.fr (A. Citanna), ostrovsky@gsb.stanford.edu (M. Ostrovsky).

0022-0531/$ – see front matter
doi:10.1016/j.jet.2009.07.004

1. Introduction

The literature on two-sided matching, starting with Gale and Shapley [7], 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 [15], 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 of 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.

When agents are uncertain about the value of their potential matches, they may infer additional information from observing the matches of others allowing them to revise their evaluations and, possibly, their own optimal actions. Indeed, top law schools tend not to hire a new professor until he or she has received several offers from other law schools. Similarly, a college senior who can credibly convey existing job offers to a recruiter finds it easier to get a new offer. Likewise, 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 the college-student market in which college c's estimate of the value of student s depends on the information of college c'. In such environments, the matching process depends on the different agents' information about the state of the world. We ask if there exist stable matching mechanisms, i.e., mechanisms that elicit information truthfully and are immune to 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. For example, 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 [10] 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. Chade, Lewis, and Smith [3] and Nagypal [14] study the behavior of students in the application process when they are uncertain about their own qualities and applications are costly. Chade [2] 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 [9] 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 due to Holmstrom and Myerson [8] and posterior efficient collective choice due to Forges [6] 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 colleges, stable mechanisms may not exist.[3] 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. Indeed, when student preferences over colleges are commonly known but diverse, a stable mechanism in general does not exist, 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, the existence of stable mechanisms 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 mediator publicly announces the profile of matches after eliciting reports from the colleges while weak stability corresponds to the case where the mediator communicates privately to each agent their own match.

We demonstrate the non-existence of stable mechanisms via non-degenerate examples. We demonstrate existence constructively via a serial dictatorship algorithm under which the match for any college is determined based on the information held only by that college and the colleges 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 colleges to the higher-ranked ones. When each college observes the entire profile of matches, a higher-ranked college will sometimes be able to infer enough information about the signals of lower-ranked colleges to object, precluding the existence of strongly stable mechanisms. But if each college observes only its own match, it is possible to construct serial dictatorship mechanisms that manage to elicit truth-telling while at the same time ensuring that the match of any college is independent of information held by lower-ranked colleges. Such an existence result obtains regardless of the preferences of the colleges 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 college only observes the realized matches of all colleges higher, but not lower, than itself in the student rankings.

We also 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 colleges that are considered better by students making offers earlier. This process is similar to the observed behavior in some labor markets, such as the 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. Section 4

[3] Roth [15] shows that with uncertainty about preferences on both sides of a matching market, stable incentive-compatible mechanisms do not generally exist even in a private-value setting. Since private values are a special case of interdependent values, Roth's non-existence results carry over, in essence, to an interdependent-value setting with two-sided uncertainty. Hence, the restriction to one-sided uncertainty is not only reasonable in our setting, since information about colleges is mostly public, but also necessary for our positive existence results.

[4] Wilson [17] discusses the desirability of creating mechanisms that are as independent of the details of the environment as possible.

presents our results on the existence and non-existence of stable mechanisms in various settings. Section 5 discusses indirect mechanisms in our setting. Section 6 concludes.

2. Model

Consider a many-to-one matching market between colleges and students. The set of students is denoted by $\mathbf{S} = \{1, \ldots, S\}$ with typical element s and the set of colleges is denoted by $\mathbf{C} = \{1, \ldots, C\}$ with typical element c. Let $k_c \geqslant 1$ be the capacity of college c, i.e., the maximum number of students it can accept.

Each student s has some quality $q_s \in \mathbf{Q}_s$, where $\mathbf{Q}_s$ is a finite set. Quality q_s is not observed by any agent in the economy (including student s). It can encompass a variety of student characteristics and as we explain below different colleges may have different preferences over student qualities. Let $\mathbf{Q} = \times_s \mathbf{Q}_s$ be the set of all quality profiles with elements $q = (q_s)_{s\in\mathbf{S}}$.

Colleges do not know the quality q_s of any student s but each college c receives a private signal $x_{c,s} \in \mathbf{X}_{c,s}$, where $\mathbf{X}_{c,s}$ is also a finite set. Let $x_s = (x_{c,s})$ be the vector of signals associated with each student s and let $x_c = (x_{c,s})$ be the vector of signals received by each college c. Let $x \in \mathbf{X} = \times_{c,s}\mathbf{X}_{c,s}$ be the vector of signals received by all colleges about all students and let $x_{-c} \in \mathbf{X}_{-c}$ be the vector of signals received by colleges other than c. Let Pr be the joint probability distribution over signals and qualities, which we assume to be strictly positive for all x, q.[5]

Our formal definition of a many-to-one matching follows [16]. Let an unordered family of elements of $\mathbf{S} \cup \mathbf{C}$ be a collection of its elements, not necessarily distinct, in which the order is immaterial. A *matching* m is a function from $\mathbf{S} \cup \mathbf{C}$ into the set of unordered families of elements of $\mathbf{S} \cup \mathbf{C}$ such that: (i) for any student $s \in \mathbf{S}$, $|m(s)| = 1$ and $m(s) = s$ or $m(s) \in \mathbf{C}$; (ii) for any college $c \in \mathbf{C}$, $|m(c)| = k_c$ and if there are $r < k_c$ students in $m(c)$, then $m(c)$ contains $k_c - r$ copies of c; and (iii) $m(s) = c$ if and only if $s \in m(c)$. In other words, each student is matched with at most one college (remaining unmatched when $m(s) = s$), while each college c is matched with at most k_c students (having unfilled seats when $c \in m(c)$). 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 with agent $a \in \mathbf{S} \cup \mathbf{C}$ at $v_{s,a}$. We normalize the value to a student from staying unmatched to $v_{s,s} = 0$ and assume that students' preferences are strict: for any student s and pair of colleges c and c', $v_{s,c} \neq v_{s,c'}$ and $v_{s,c} \neq 0$. If $v_{s,c} < 0$, college c is unacceptable for student s; otherwise it is acceptable.

The payoffs of colleges depend on students' qualities and signals: college c derives utility $w_{c,a}(x, q)$ from matching with agent a given signals x and qualities q. Notice that this allows the values of a student to be completely common to all colleges (e.g., $\mathbf{Q}_s \subset \mathbb{R}$ and $w_{c,s}(x, q) = q_s$) or completely private (e.g., $\mathbf{X}_{c,s} \subset \mathbb{R}$ and $w_{c,s}(x, q) = x_{c,s}$), as well as to be partially interdependent. We normalize to zero the value $w_{c,c}(x, q)$ to a college from keeping a seat vacant and assume that the value of a match $m(c)$ to college c is additive, i.e., $w_{c,m(c)}(x, q) = \sum_{a\in m(c)} w_{c,a}(x, q)$.

Colleges and students have von Neumann–Morgenstern preferences. In particular, college c's expected utility from match $m(c)$ given information I is equal to $u_{c,m(c)}(I) = \sum_{x,q} w_{c,m(c)}(x, q) \Pr(x, q|I)$.

[5] Notice that we allow informational spillovers by letting signals and qualities to be correlated not only across colleges but also across students. The full support assumption is made only to simplify the presentation of the 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, colleges will update their beliefs about student qualities differently under different mechanisms (and under different amounts of information). As a result, a particular matching may be stable under one mechanism but unstable under another. Consequently, 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 colleges.

Of course, this is not the first paper to discuss stable matching mechanisms (see, e.g., [15]). However, in [15] and other related papers the focus is on whether there is an incentive-compatible mechanism that will produce a stable matching when each agent's preferences over his matches are independent of the mechanism and the information revealed by it. In contrast, in our setting, the agent's preferences over his matches may depend on the revealed information.

3.1. Definitions

We begin by formalizing the concept of a "matching mechanism." We define it as a centralized direct revelation mechanism, in which colleges report their signals to a mediator and the mediator then proposes who should be matched with whom. More formally, a *direct revelation matching mechanism* μ is a function from the set $\hat{\mathbf{X}} \times [0,1]$ of reported signal profiles and draws of a random variable ω to the set $\mathbf{M}$ of matchings. The presence of ω captures the fact that, as a function of the signals reported by colleges, the final matching can be stochastic, i.e., the mediator can randomize. Without loss of generality, we assume that ω is distributed uniformly on $[0,1]$.

We now turn to our concept of 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 any of his matches and no college-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, but with two important differences: first, we define stability on matching mechanisms, and second, the definition takes into account how much information will be available to the colleges at the time of rematching. The amount of information available to the colleges at the time of rematching will be a crucial determinant of the stability of a matching mechanism.

We capture the information available to the agents, in particular to colleges, at the time of rematching by a vector of *message functions* $\mathbf{z} = (z_a)_{a \in \mathbf{S} \cup \mathbf{C}}$, where each z_a is a function from the set $\hat{\mathbf{X}} \times [0,1]$ of reported signal profiles and draws of ω to a finite set $\mathbf{Z}_a$ of possible messages received by agent a. We denote by $z_a(\hat{x}, \omega)$ the message received by agent a after colleges reported signals $\hat{x}$ and the mechanism drew random variable ω for the randomization. Note that while z_a's are deterministic functions, our setting allows for noisy information revelation about the variables of direct interest to the agents: the reported signals and the resulting matching. For instance, for some vector of reports, mechanism μ may output a deterministic matching and at the same time the message z_a received by agent a may be different for $\omega \leqslant 0.5$ and $\omega > 0.5$, thus resulting in noisy information about the produced matching.

We assume that at the very least, message z_a must reveal to agent a its own match $m(a)$. That is, there do not exist vectors of reports $\hat{x}^1$ and $\hat{x}^2$ and draws of random variable ω^1 and ω^2 such

that $z_a(\hat{x}^1, \omega^1) = z_a(\hat{x}^2, \omega^2)$ and $\mu_a(\hat{x}^1, \omega^1) \neq \mu_a(\hat{x}^2, \omega^2)$, where $\mu_a(\cdot)$ denotes the match of agent $a \in \mathbf{S} \cup \mathbf{C}$. Message z_a can also reveal the entire matching m and all reported signals $\hat{x}$, as well as information about the draw of ω. Crucially, different specifications of information structure $\mathbf{z} = (z_a)_{a \in \mathbf{S} \cup \mathbf{C}}$ at the time of rematching will correspond to different notions of stability.

Given matching mechanism μ and information structure $\mathbf{z}$, we define extensive form game $\Gamma(\mu; \mathbf{z})$ as follows. The players are all students in $\mathbf{S}$ and all colleges in $\mathbf{C}$. The game consists of the following stages:

1. Nature selects qualities q and signals x according to the commonly known distribution Pr and communicates x_c to each college c.
2. Colleges simultaneously send their reports, $\hat{x}_c$, to the mechanism.
3. The mediator draws ω at random and generates matching $m = \mu(\hat{x}, \omega)$.
4. Each agent observes his message z_a.
5. At this stage, colleges simultaneously choose one of several actions. First, college c can do nothing. Second, if college c is matched to at least one student, it can pick one of them, say $s \in m(c)$, and unilaterally drop him from its match. Finally, if college c and some student s are unmatched, i.e., $s \notin m(c)$, college c can choose to make a rematching offer to this student. If it decides to do that, and it does not have any vacant seats (i.e., the number of students in $m(c)$ is equal to the colleges's capacity k_c), it also needs to decide which student $s' \in m(c)$ it will drop if student s accepts the offer.
6. At the last stage, any student s who has one or more rematching offers can choose to accept at most one of them, possibly rejecting all of them and keeping his assigned match.

If in stage 5 college c chose to do nothing, we say that it accepted its match. If it chose to drop a student, or if it made a rematching offer that was subsequently accepted, we say that it successfully objected to the match.

The ex-post payoffs in game $\Gamma(\mu; \mathbf{z})$ are as follows. Any student s who accepted a rematching offer from college c is matched to that college and therefore gets payoff $v_{s,c}$. Of the remaining students, any student s who was unmatched in m, was unilaterally dropped by its assigned college in stage 5, or was dropped by its assigned college because of that college's successful rematching offer to some other student stays unmatched and therefore gets payoff $v_{s,s} = 0$. All remaining students s are matched to their assigned colleges, $m(s)$, and get payoffs $v_{s,m(s)}$. The ex-post payoff of college c is its utility from the final match it receives after stage 6, $w_{c,m'(c)}(x, q)$. Note that the college's final match $m'(c)$ can differ from its assigned match $m(c)$ both because of its own actions (dropping a student or successfully making a rematching offer) and because of the actions of other colleges (college c' making a successful rematching offer to one of the students in $m(c)$ and thus stealing that student from college c).

We can now define the concept of stability.

Definition 1. Matching mechanism μ is stable under information structure $\mathbf{z}$ if it never matches a student to an unacceptable college and there exists a perfect Bayesian equilibrium of game $\Gamma(\mu; \mathbf{z})$ in which all colleges report their signals truthfully and each college accepts its assigned match on the equilibrium path.

In essence, our notion of stability embeds three requirements. The first is incentive compatibility: colleges do not have an incentive to lie about their signals. The second is no rematching after truthful information revelation: no college can report its signal truthfully and then profitably

rematch after the mechanism proposes a particular matching and messages z are observed (assuming other colleges report signals truthfully as well). The third requirement is a combination of the first two: no college 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.

Rematching-proofness has already been used by Ma [11] to refine Nash equilibria in matching markets. Our stability notion, however, differs from Ma's notion of rematching-proof equilibrium in two important respects: first, we do not allow a college and a student to *jointly* deviate in the reporting stage, but only in the rematching stage, whereas Ma also allows them to coordinate their reports; second, Ma considers the setting of Roth [15], in which players have private values, whereas in our settings colleges are uncertain about their preferences over students and therefore use information on the matching outcome to update beliefs at the posterior stage.

More broadly, our restrictions are similar to the notions of "truthful" and "obedient" behavior by agents in the "direct coordination" mechanisms for generalized principal-agent problems with private information and private decisions, studied by Myerson [13]. In particular, the possibility of anticipated rematching implies that the mediator does not have the final say about the realized allocation, and hence incentive compatibility is not sufficient for stability. For example, a uniform random matching mechanism that ignores all reports and assigns every student to every college with equal probabilities is not necessarily stable, even though there is no incentive for any college to misreport its signal. As a result, the basic version of the revelation principle [12] does not apply to our setting, since it is only concerned with the reporting of signals. Technically, the more general revelation principle of [13] does not apply either, because that paper only considers one-stage games, whereas in our setting colleges first propose rematching offers and then students choose whether or not to accept them. This difference, however, turns out to be very minor, and as we argue in Section 5.2, using essentially the techniques of [13], restricting attention to direct revelation mechanisms is without loss of generality.

The rematching stages of Γ are also quite specific and deserve comment. Notice first that if μ is stable under information structure $\mathbf{z}$, then in the corresponding equilibrium each student s must end up matched with the $m(s)$ assigned to it by μ on the path of play. This is because μ never matches a student to an unacceptable college and, when no college objects to its assigned match, either a student has exactly one acceptable offer from college $m(s) \in \mathbf{C}$ or the student has no available offers and $m(s) = s$. Notice next that, when μ is stable, no college expects any other college to object to its assigned match, at least on the path of play. In effect, our notion of stability allows a single round of rematching and requires no agent to object to its assignment when it expects no other agents to object. As alternatives, we could have allowed multiple colleges to rematch in an exogenous or endogenous sequence 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 chance at rematching.[6]

Finally, Definition 1 contains in it a number of distinct notions of stability that differ with respect to information structure $\mathbf{z}$. The case where each z_a reveals to agent a only its own match $m(a)$ corresponds to the notion of *weak* stability. In contrast, we use the term *strong* stability to refer to the case where in addition z_a reveals to each agent a the matches $m(a')$ of all other agents a'. One could also consider the notion of stability under even more detailed information: z_a could reveal not only the entire profile of matches m, but also all reports $\hat{x}$ submitted by the agents to the mediator. Just like in [15], such a notion would be similar to ex post stability in the sense that a mechanism stable under this information structure will produce matchings that are stable in the classical sense for each profile of signals x. Notice that in our setting the information communicated to students does not play a role, as long as each student knows his own match.

4. Results

In this section we investigate how stability depends on the information structure $\mathbf{z}$ and other features of the matching environment. As a background, our first result makes precise the intuition that more information makes achieving stability more difficult. We say that information structure $\mathbf{z}$ is *coarser* than information structure $\mathbf{z}'$ if for each a, for every pair of reports $\hat{x}^1$ and $\hat{x}^2$ and every pair of draws ω^1 and ω^2, $z_a(\hat{x}^1, \omega^1) \neq z_a(\hat{x}^2, \omega^2)$ implies $z'_a(\hat{x}^1, \omega^1) \neq z'_a(\hat{x}^2, \omega^2)$. Equivalently, there exist mappings $\zeta_a : \mathbf{Z}'_{\mathbf{a}} \to \mathbf{Z}_{\mathbf{a}}$ such that $z_a(\hat{x}, \omega) \equiv \zeta_a(z'_a(\hat{x}, \omega))$. In words, this definition says that everything about the inputs to (and thus the outputs of) the mechanism that agent a knows under information structure $\mathbf{z}$ he also knows under $\mathbf{z}'$.[7] For instance, the information structure corresponding to weak stability (where agents observe only their own matches) is coarser than the one corresponding to strong stability (where they observe the entire profile of matches). Similarly, the information structure that corresponds to strong stability is coarser than the information structure that reveals to the agents both the entire matching outcome and the colleges' reports to the mechanism.

Theorem 1. *If μ is stable under some information structure it is also stable for every coarser information structure.*

Proof. See Appendix A. □

While the proof contains some important technical details, the key intuition behind the result is straightforward. Consider the same mechanism μ in two information regimes, one coarser than the other. If a college has an incentive to rematch or drop its assigned match after observing some

[6] This approach follows in spirit that of the original definition of stability in two-sided one-to-one matching markets: 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.

[7] This definition is similar in spirit to Aumann's [1] definition of finer and coarser partitions, but there is an important difference: while under Aumann's definition, information partitions refer to the agents' knowledge about the underlying states of the world, in our case the information is about the signals reported by the colleges to the mechanism and the random variable drawn by the mechanism; beyond that, no information about the true underlying state of the world prior to the mechanism's execution (the signals received by the colleges or the qualities of the students) is revealed. If colleges were forced to mechanically report their signals to the mechanism, then our definition of finer and coarser information would be equivalent to Aumann's definition.

information in the coarser information regime, then such an action must also be profitable after observing at least one possible signal in the finer information regime, making the mechanism unstable in the finer information regime as well.

Corollary 2. *Any strongly stable mechanism μ is weakly stable.*

4.1. 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 colleges are identical. We show this by presenting an example of a whole class of matching markets for which no such mechanism exists: markets with two colleges, three students, and two quality levels and signals. By Theorem 1, the same example shows that stable matching mechanisms do not generally exist when agents observe not only the entire matching produced by the mechanism but also the reports of all colleges.

Example 1. Consider a market with two colleges, 1 and 2, each with a capacity $k_c = 1$ 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 $1 - \theta$ independently across students. If the student is high type, each college gets signal H about him with probability p_h, and signal L with probability $1 - p_h$. If the student is low type, each college 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 college 1 more than college 2: for any s, $v_{s,1} > v_{s,2} > 0$. Colleges' preferences are identical, with $w_{i,s}(x, q) = q_s$. Note that the information structure is such that a student with two H signals is in expectation strictly more valuable to a college than a student with one H and one L signal, who in turn is strictly more valuable than a student with two L signals. Note also that every student is acceptable.

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

Proof. 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 college 1 gives the less desirable college 2 incentives to conceal its information in the reporting stage in the hopes of obtaining the student later, perhaps through rematching. On the other hand, assigning this student to college 2 gives incentives to college 1 to steal college 2's assigned match after observing the realized match. The possibility of assigning the student stochastically does not help achieve stability, as we show below.

Specifically, suppose μ is a strongly stable mechanism for the market described in Example 1. We establish the claim in four steps:

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

To see this notice that college 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. So for college 1 to be willing not to rematch, the probability of benefiting from rematching has to be zero. Therefore,

Table 1
Value of college 2's final match as a function of college 1's signals (Step 4).

Signals x_1	HH students	HL students	Value of final match for college 2
HHH	s_1, s_2	s_3	HH
HHL	s_1, s_2	–	HH
HLH	s_1	s_2, s_3	HL with prob. $1/2$, HH with prob. $1/2$
LHH	s_2	s_1, s_3	HL with prob. $1/2$, HH with prob. $1/2$
HLL	s_1	s_2	at least HL
LHL	s_2	s_1	at least HL
LLH	–	s_1, s_2, s_3	HL
LLL	–	s_1, s_2	HL

μ cannot match college 1 with, say, s_2 if student s_1 is strictly better with positive probability, i.e., when college 2's signal about s_1 is H or college 2's signal about s_2 is L.

Step 2. If college 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 college 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: college 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.

Step 3. Suppose college 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 college 1, whenever there is such a student.

To see this notice that college 1 can guarantee itself a student with two H signals whenever there is one. Indeed, college 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 college 2 has observed H about that student who is therefore an HH one. If college 1 gets matched to s_3, it should rematch with the student matched to college 2: by Step 2, college 2's signal about that student is H, unless college 2's signals about both s_1 and s_2 are L, in which case there are no HH students anyway.

Step 4. Now suppose college 2 observes two Hs (say, s_1 and s_2) and one L (s_3): $x_2 = (H, H, L)$. Then it is not optimal for college 2 to report its signals truthfully and accept its assigned match.

Indeed, by Steps 2 and 3 it is easy to see that under truthful reporting with no rematching, college 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 college 2 can strictly increase its payoff with the following deviation: it should first "invert" the signals reporting $\hat{x}_2 = (L, L, H)$; then (i) if college 2 is matched to s_3, it should rematch with the unassigned student, (ii) if college 1 is matched to s_3, college 2 should rematch with the unassigned student with probability $1/2$ and keep its assigned match with probability $1/2$, and (iii) if s_3 remains unassigned, college 2 should not rematch. Table 1 lists the HH and HL students and the value (profile of signals) of the final match for college 2 under this deviation strategy for each possible realization of college 1's signals.

To understand the table, notice first that under the deviation strategy college 2 always ends up matched with either s_1 or s_2. It follows that college 2 always ends up matched to at least a

HL student and, when there are two HH students, it ends up matched with one of them, just as it does under truthful reporting and no rematching. However the deviation strategy also allows college 2 to get matched with the only HH student with strictly positive probability when the signal of college 1 is either HLH or LHH. In such situations, by Step 3, college 1 must get matched with s_3 so that college 2 finally matches with s_1 or s_2 with equal probabilities under the deviation strategy. It follows that college 2's deviation strategy is at least as good as the strategy of truthful reporting for each possible realization of x_1 and strictly better for some. Therefore, this is a profitable deviation also in expectation, conditional on $x_2 = HHL$. □

Notice that in the argument above we make crucial use of the possibility of anticipated rematching, i.e., the possibility that college 2 may misreport its signal and rematch after observing the profile of matches. The argument also depends on the binary nature of signals and the presence of three students (for instance in Step 2). We have constructed another example that does not depend on these features. It relies instead on the presence of an arbitrarily small private value component in the colleges' preferences. The example is available from the authors upon request.

Also, while Example 1 shows that strongly stable mechanisms do not generally exist, it does not prove that they cannot exist for specific environments. We have constructed an example of a strongly stable mechanism in which information is transmitted and is used by the mediator in a non-trivial manner. This example is also available upon request.

Finally, while Example 1 is constructed for a matching market with only three students and two colleges, it can be embedded in a larger matching market for which the non-existence result will also hold. One simple way to do this is to assume that no other student is acceptable to the two colleges in the example under any combination of signals, that all other colleges are less desirable to the students than these two, and that all other colleges receive completely uninformative signals about the three students in the example.

4.2. Weak stability with heterogeneous student preferences

As the results of the previous section show, there does not in general 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. We show that when students' preferences over colleges differ, even a weakly stable matching mechanism may not exist. As we show in the next section, however, when students have identical preferences such a mechanism does exist.

The intuition behind the next example is simple. A college that is low in the ranking of its most preferred student (say, student 1) may lie in order to mislead a college 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 college if the college thinks that student is best, and since student rankings are not identical, there is room to change the available set of matches through misreporting.

Example 2. Consider an environment with two colleges, 1 and 2, each with a capacity $k_c = 1$, and two students, s_1 and s_2. Students' preferences $v_{s,c}$ 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 $1/2$ each. College 1's signal about q_2 is uninformative. College 2's signal about

q_2 is perfectly informative; it is equal to q_2 with probability 1. Colleges' preferences are summarized as follows:

$$w_{1,s_1}(x,q) = 1,$$
$$w_{1,s_2}(x,q) = -q_2,$$
$$w_{2,s_1}(x,q) = 3,$$
$$w_{2,s_2}(x,q) = q_2.$$

Claim 4. *The matching market described above does not have a weakly stable mechanism.*

Proof. We identify x with the signal x_{2,s_2} of college 2 about student s_2 since x_1 and x_{2,s_1} are both uninformative. Slightly abusing notation, let $\mu(m|\hat{x})$ denote the probability of the mechanism generating matching m conditional on reports $\hat{x}$. There are two possible signals reported to the mechanism, $x = 2$ and $x = -2$. In what follows, we summarize a matching by specifying the matches for the two colleges, $(m(1), m(2))$. There are seven such matches that the mechanism can produce: (s_1, s_2), $(s_1, 2)$; (s_2, s_1), $(s_2, 2)$; $(1, s_1)$, $(1, s_2)$, and $(1, 2)$. We identify the necessary properties of a stable matching mechanism μ in order to show that no such mechanism exists.

Since college 1's payoff from matching with student s_1 is always equal to 1 and its payoff from staying unmatched is equal to zero and since student s_1 prefers college 1 to college 2, a stable mechanism cannot leave college 1 unmatched, i.e., any stable mechanism must have $\mu(m|x) = 0$ if $m(1) = 1$, for any x. This leaves the four possible matches (s_1, s_2), $(s_1, 2)$, (s_2, s_1), and $(s_2, 2)$ that the mechanism can output with positive probability.

Likewise, since college 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 college 2 to staying unmatched, a stable mechanism cannot leave both college 2 and student s_1 unmatched, i.e., any stable mechanism must have $\mu(m|x) = 0$ if $m(2) = 2$ and $m(s_1) = s_1$. This leaves the three possible matches (s_1, s_2), $(s_1, 2)$, and (s_2, s_1) that the mechanism can output with positive probability.

Furthermore, since $w_{2,s_2}(x, q_{s_2}) = q_{s_2} = x$ and $w_{2,s_1}(x, q_{s_1}) = 3$, any stable mechanism must have $\mu(m|x) = 0$ if $m(2) = s_2$ and $x = -2$ or if $m(2) = 2$ and $x = 2$. For if not, in the first case college 2 will object by dropping its assigned student s_2 and staying unmatched, while in the second case college 2 will object by making an offer to student s_2 that the student will accept regardless of its assigned match. In other words, the mechanism can output the match (s_1, s_2) with positive probability only if $x = 2$ and it can output the match $(s_1, 2)$ with positive probability only if $x = -2$.

If the mechanism outputs the match $(s_1, 2)$ (in particular, leaving s_2 unmatched) with strictly positive probability when $x = -2$, then upon observing $m(1) = s_1$ college 1 could strictly improve its expected payoff by making a rematching offer to student s_2: such an offer will be accepted if student s_2 is unmatched and rejected otherwise. In the latter case, college 1 will keep its assigned match s_1. Since college 1 prefers s_2 to s_1 only when $x = -2$, and since s_2's acceptance of the rematching offer reveals the state to college 1, such a rematching offer benefits college 1. This is the key step in the proof where we use the fact that college 1 infers some additional information about the state of the world from a student's acceptance decision of a rematching offer. It allows us to conclude that when $x = -2$, a stable matching mechanism must output the match (s_2, s_1) with probability 1.

On the other hand, when $x = 2$, we know already that a stable mechanism can either output the same matching (s_2, s_1) or the matching (s_1, s_2). But if the latter matching occurs with strictly positive probability when $x = 2$, college 2 would have a strict incentive to report $\hat{x} = -2$ whenever its true signal is $x = 2$. Hence, the only remaining candidate for a stable mechanism is the deterministic mechanism μ that always outputs matching (s_2, s_1) for any report of college 2. However, this mechanism also cannot be stable since college 1 can make a rematching offer to s_1 that will always be accepted, yielding college 1 a greater payoff than the expected payoff of zero from always matching with s_2. □

Notice that in the proof of Claim 4 we derive the necessary properties of a stable matching mechanism by considering only the possibility of rematching after truth-telling and the possibility of misreporting private signals without any subsequent rematching. The argument does not rely on the deviations involving anticipated rematching. It follows that the result would continue to hold even under weaker definitions of stability than the one we use.

4.3. Weak stability with homogeneous student preferences

Section 4.2 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 college is the most desirable, the second most desirable, and so on. Formally,

$$v_{s,c} = v_{s',c} \equiv v_c \quad \text{for any students } s \text{ and } s' \text{ and college } c.$$

College 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 generally stable mechanism even with identical preferences of students over colleges.

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 c_1 be the top-ranked college. Compute each student's value to college c_1 based on the college's own signal, and assign to college c_1 the $r_{c_1} \leqslant k_{c_1}$ students with the highest value (obtaining $m(c_1)$), where r_{c_1} is equal to k_{c_1} if there are at least k_{c_1} students acceptable to c_1 based on its information and r_{c_1} is equal to the number of students acceptable to c_1 if that number is less than k_{c_1}. Break ties via uniform randomization. For the second-ranked college, c_2, compute the value of each remaining student based on the college's own signal, as well as on observing $m(c_1)$. Assign to college c_2 the $r_{c_2} \leqslant k_{c_2}$ students among those remaining with the highest value (obtaining $m(c_2)$). For the third-ranked college, c_3, compute the value of each remaining student based on college c_3's own signal and on observing $m(c_1)$ and $m(c_2)$. Proceeding in this fashion, we assign a match to each acceptable college (i.e., each college c with $v_c > 0$); all unacceptable colleges remain unmatched. We denote this mechanism as μ^{SD}.[8]

[8] Note that there are other ways to implement analogues of serial dictatorship in this environment, requiring less information. For instance, in assigning matches to low-ranked colleges, the mechanism can condition only on which set of students was matched to higher-ranked colleges without using knowledge of which student went to which particular college. Theorem 5 would continue to hold under this formulation, although the resulting matching would not in general be the same. Note also that since our specification of college preferences includes the case of private values, the result carries over to an environment similar to that of [15], except that matching is many-to-one instead of one-to-one and that the preferences of students are restricted to be identical. Since in that environment preferences of one college do not

Theorem 5. *Under homogeneous student preferences, mechanism μ^{SD} is weakly stable.*

Proof. See Appendix A. □

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

The arguments underlying Theorem 5 also make precise the degree of transparency under which stable mechanisms generally exist. In particular, if colleges observe the profile of matches for lower-ranked colleges, a higher-ranked college will typically be able to infer information held by lower-ranked colleges from observing the matching outcome of lower-ranked colleges. It follows that existence is guaranteed in general, if each college observes at most the matches of all higher (but not lower) ranked colleges.

Note also that in many plausible situations the information held by lower-ranked colleges may not be relevant for the evaluation of students by higher-ranked colleges. For instance, higher-ranked graduate schools may be better at evaluating the suitability of a student for graduate school. In such cases, existence is not an issue even when each college observes the entire profile of matches including those of lower-ranked colleges. In this sense, the proof of Theorem 5 also identifies economically interesting environments for which strongly stable mechanisms will exist.

5. Indirect mechanisms

So far, the paper has focused on direct mechanisms, in which colleges observe signals about students, report them to the mediator, and then choose whether to accept the mediator's proposed match or to deviate. In many settings, however, indirect or decentralized mechanisms may have useful properties, and we now turn out attention to such mechanisms. First, we show that the serial dictatorship mechanism presented in Section 4.3 can be implemented as a detail-free, decentralized game. Second, by using the revelation principle, we show that the restriction to direct mechanisms is without loss of generality.

5.1. Detail-free implementation of serial dictatorship

In Theorem 5 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 colleges can receive about students, compute expected payoffs to colleges 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 common ranking of colleges by the students; if this ranking is not known to the mediator (but he knows

depend on the information of others, there is only one way to implement serial dictatorship. Moreover, there is a unique stable matching (the best college is matched to its preferred set of students who find it acceptable; the second best college is matched to its preferred subset of remaining students who find it acceptable, and so on), and serial dictatorship results in that stable matching.

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.[9]

The game proceeds in rounds, $t = 1, 2, \ldots, C$. In each round t, college c_t is allowed to make an offer to at most k_{c_t} students among the available ones; it is also allowed to make no offers (recall that c_1 is the most desirable college, c_2 is the second most desirable, and so on). The students to whom offers are made can accept them or reject. After the round is over, colleges $c_{t+1}, \ldots, c_C$ are told which matches, if any, were 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 college c, having observed what happened in the previous stages, updates its beliefs on the signals of other colleges using Bayes' rule whenever possible. If what a college 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 colleges c_1 through c_t have a positive probability according to its priors and updates the beliefs based only on the matches of colleges c_1 through c_t, ignoring the subsequent matches. Any other internally consistent way of updating beliefs after zero-probability events would also work. Having computed these beliefs, college c makes offers to up k_c of the available students with the highest expected payoffs, provided those payoffs are positive (if there are fewer than k_c acceptable students, the college makes offers to all of them; otherwise, it makes offers to the k_c students with the highest expected payoffs). Students always accept offers from acceptable colleges and reject offers from unacceptable ones.

It is easy to check that no college 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.

5.2. The revelation principle

Most of the results in this paper use the definitions of stability for centralized direct revelation mechanisms. The only exception is Section 5.1, 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 in generalized principal-agent problems [13], 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. 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 G 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

[9] 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.

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 game G', which extends game G with a single simultaneous rematching stage (just like in the direct revelation mechanism case) has 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 G and profile of strategies σ in a stable way.

Now suppose that there is a stable indirect mechanism (game G with a corresponding profile of strategies σ) for a particular matching market. We argue below that there also exists a strongly stable direct revelation mechanism for this market.

To see this, consider the following direct revelation mechanism μ *under information structure* $\mathbf{z}$: (i) colleges observe their signals and report them to the central mediator; (ii) the mediator, in the background, runs game G in accordance with these reports and strategies σ; (iii) the mediator outputs to the colleges and students the final matching *and also all the extra information that they would have observed along the way if they played the game themselves*.[10]

Note that mechanism μ is stable under information structure $\mathbf{z}$, since any feasible 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 can "imitate" someone of a different type.

Now consider the same direct revelation mechanism μ under information structure $\mathbf{z}_{strong}$ that reports to the colleges and students only the final matching, without all the additional information that they would have observed along the path of play in game G. Information structure $\mathbf{z}_{strong}$ is a coarsening of information structure $\mathbf{z}$ and so by Theorem 1, mechanism μ is stable under $\mathbf{z}_{strong}$, i.e., μ is strongly stable.

6. 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 colleges are ranked higher precisely because they are better able to identify suitable students. In such cases, the information held by lower-ranked colleges is unlikely to matter for evaluations by higher-ranked colleges, 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)

[10] Since the extensive form game is finite, there is only a finite amount of information to convey, and if players randomize along the way, this randomization also takes place at a finite number of nodes and involves a finite number of possible moves. Thus, the uncertainty in this mechanism can be summarized by one random variable ω, distributed uniformly on $[0, 1]$. To give a concrete example, suppose there are two colleges, with college 1 choosing action a with probability, say, $3/4$ and action a' with probability $1/4$, and college 2 choosing action b with probability $1/3$ and action b' with probability $2/3$. Each college knows its own action but not the action of the other player. This setting can be captured as follows. Let ω be the random draw from $U[0, 1]$. If $\omega \leqslant 3/4$, μ plays a for college 1; otherwise, it plays a'. If $\omega \leqslant 3/4 \cdot 1/3$ or $\omega \in (3/4, 3/4 + 1/4 \cdot 1/3]$, then μ plays b for college 2; otherwise, it plays b'. Information structure $\mathbf{z}$ simply reports to each college the action that the mechanism took on its behalf (e.g., for $\omega \leqslant 3/4 \cdot 1/3$, $z_1(\omega) = a$ and $z_2(\omega) = b$).

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 colleges 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.

Since our results show that in many settings stable mechanisms do not exist, a natural question is which outcomes can be sustained if the mechanism designer can have colleges sign binding contracts not to renege in the rematching stage, but still needs to provide them with enough incentives ex ante to participate in the mechanism rather than bypass it altogether. Indeed, an agreement not to raise objections to the mechanism's output may be self-enforcing in a dynamic setting in which a college is punished for reneging on the agreement by exclusion from future matching markets. The analysis of such effects is beyond the scope of this paper, but is an interesting avenue for future research.

Finally, we focused on the notion of pairwise stability, allowing colleges 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 [16] and Crawford [4].[11] The study of stability in matching markets under various notions of group stability is another possible area for future research.

Acknowledgments

We would like to thank the Associate Editor, two anonymous referees, Atila Abdulkadiroglu, Dirk Bergemann, Francoise Forges, Drew Fudenberg, Christine Jolls, Emir Kamenica, 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 helpful comments and suggestions. This paper stems from and adds to results that previously circulated in two separate papers. Parts of the paper were completed while the second author was visiting Columbia Business School and the third author was visiting the Hoover Institution at Stanford. The first and the second authors acknowledge the financial support of NSF grant #0617850.

Appendix A. Proofs

Proof of Theorem 1. Suppose mechanism μ is stable under information structure $\mathbf{z}'$ and consider a coarser information structure $\mathbf{z}$. We want to show that μ is also stable under $\mathbf{z}$.

If student s is matched to an unacceptable college c by μ under $\mathbf{z}$, then $v_{s,c} < 0$. Since the student's payoff $v_{s,c}$ is independent of the information structure, then the student would also have found c unacceptable under $\mathbf{z}'$. This proves that mechanism μ satisfies the first part of the definition of stability under $\mathbf{z}$—it never matches a student to an unacceptable college.

The proof of the second part is more involved. We will show the statement by contrapositive, that is, if truthful reporting and the agents' acceptance of their assigned match is not a PBE of the game $\Gamma(\mu; \mathbf{z})$, then it cannot be a PBE of $\Gamma(\mu; \mathbf{z}')$. It is straightforward to show that no student

[11] For example, [16, 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 [4, p. 394].

can improve his payoff by deviating. Then, we need to check three forms of deviations, only for colleges: deviating in the rematching stage (dropping a student or making a rematching offer), deviating in the reporting stage (misreporting the signal), or doing both.

We first make two observations. Consider college c. First, note that any rematching-stage information set for c is fully specified by its signal x_c, its report $\hat{x}_c$, and the observed message (z_c or z'_c, depending on the information structure). Crucially, given μ, x_c, and $\hat{x}_c$, the rematching-stage information of c under $\mathbf{z}'$ is always finer than under $\mathbf{z}$.

Second, by the full-support assumption on signals and qualities, the rematching-stage beliefs of c in both games (if c believes that others are reporting their signals truthfully) are uniquely pinned down by Bayes' rule, and for any set of reports $\hat{x}$ and any message given these reports, the outcome observed by c is consistent with the assumption of truth-telling by other players. In other words, c will never observe zero-probability events if he believes that other colleges are reporting their signals truthfully.

We now check that if any of the three forms of deviations is profitable in $\Gamma(\mu;\mathbf{z})$, then it is also profitable in $\Gamma(\mu;\mathbf{z}')$.

Suppose after a particular realization of signals x with $x_c = x_c^*$, a resulting match m, and a message z_c under $\mathbf{z}$, college c finds it strictly profitable to drop some student s at stage 5, i.e.,

$$\begin{aligned} u_{c,s}\big(x_c^*, z_c\big) &= \sum_{x_{-c},q} w_{c,s}\big(x_c^*, x_{-c}, q\big) P\big(x_{-c}, q \,|\, x_c^*, z_c\big) \\ &= \sum_{x_{-c},q} w_{c,s}\big(x_c^*, x_{-c}, q\big) \frac{P(z_c|x_c^*, x_{-c}, q)\Pr(x_{-c}, q|x_c^*)}{P(z_c|x_c^*)} < 0 = u_{c,c}, \end{aligned} \tag{1}$$

where $P(\cdot|\cdot)$ is the conditional probability operator, given underlying distribution Pr of qualities and signals, the assumption of truth-telling by other players, mechanism μ, and information structure $\mathbf{z}$. Let $P'(\cdot|\cdot)$ be an analogous conditional probability operator for the finer information structure $\mathbf{z}'$. Then $P(z_c|x_c^*, x_{-c}, q) = \sum_{z'_c|\zeta_c(z'_c)=z_c} P'(z'_c|x_c^*, x_{-c}, q)$. Substituting,

$$\begin{aligned} &\sum_{x_{-c},q} w_{c,s}\big(x_c^*, x_{-c}, q\big) \sum_{z'_c|\zeta_c(z'_c)=z_c} P'\big(z'_c|x_c^*, x_{-c}, q\big) \Pr\big(x_{-c}, q|x_c^*\big) \\ &\quad = \sum_{z'_c|\zeta_c(z'_c)=z_c} \sum_{x_{-c},q} w_{c,s}\big(x_c^*, x_{-c}, q\big) P'\big(z'_c|x_c^*, x_{-c}, q\big) \Pr\big(x_{-c}, q|x_c^*\big) < 0, \end{aligned} \tag{2}$$

which implies that there exists z'_c such that

$$\sum_{x_{-c},q} w_{c,s}\big(x_c^*, x_{-c}, q\big) P'\big(z'_c|x_c^*, x_{-c}, q\big) \Pr\big(x_{-c}, q|x_c^*\big) < 0. \tag{3}$$

Dividing by $P'(z'_c|x_c^*)$, we have

$$u_{c,c} = 0 > \sum_{x_{-c},q} w_{c,s}\big(x_c^*, x_{-c}, q\big) \frac{P'(z'_c|x_c^*, x_{-c}, q)\Pr(x_{-c}, q|x_c^*)}{P'(z'_c|x_c^*)} = u_{c,s}\big(x_c^*, z'_c\big), \tag{4}$$

i.e., college c would also strictly prefer to drop student s under $\mathbf{z}'$ at z'_c, at stage 5 of $\Gamma(\mu\mathbf{z}')$. That is, if μ is unstable for this reason under $\mathbf{z}$, then it is also unstable for the same reason under $\mathbf{z}'$.

Similarly, suppose that after a particular realization of signals x with $x_c = x_c^*$ and message z_c under $\mathbf{z}$, college c finds it strictly profitable to make an offer to student $s \notin m(c)$ instead of agent $a \in m(c)$, with $a = c$ or $a \neq s$. In this case, it must be the case that the probability of acceptance

by s, conditional on x_c^* and z_c, is positive and that the expected utility of c from matching with s, conditional on x_c^*, z_c, and the acceptance by s, is strictly higher than the expected utility of c from matching with a conditional on the same information. But by the same logic as in the previous case, this implies that there must exist a refinement of z_c, z_c', such that the probability of acceptance by s, conditional on x_c^* and z_c' is positive and the expected utility of c from matching with s, conditional on x_c^*, z_c', and the acceptance by s, is strictly higher than the expected utility of c from matching with a conditional on the same information. This, in turn, implies that μ is not stable under $\mathbf{z}'$.

Let us now look at the second type of deviation—misreporting its signal. Suppose c misreports its signal in $\Gamma(\mu; \mathbf{z})$. Then it will get the same distribution over matches as it would by misreporting its signal in the same way in $\Gamma(\mu; \mathbf{z}')$, because (i) mechanism μ does not depend on the information structure and (ii) in both $\Gamma(\mu; \mathbf{z})$ and $\Gamma(\mu; \mathbf{z}')$, the other colleges will accept their matches after any misreporting by college c. Part (ii) of the statement holds because of the full-support assumption over signals and qualities: any information that another college d receives is consistent, from d's point of view, with equilibrium behavior by others, and since by assumption d does not object to the match on the equilibrium path, it will not object to it after any combination of signals that it receives. But this implies that if misreporting is strictly profitable to c in game $\Gamma(\mu; \mathbf{z})$, it is also strictly profitable in $\Gamma(\mu; \mathbf{z}')$.

Finally, let us examine the deviation consisting in first misreporting its signal and then successfully objecting to its assigned match. We have already shown that the behavior of other colleges is not affected by the misreporting of college c—they always accept their match. Hence, college c can ignore the strategic behavior of other colleges when it considers its deviations, and simply view them as mechanical players who report their signals truthfully and then do nothing else; in other words, from college c's point of view, the game becomes a decision problem. Note that in this decision problem, under truthful reporting and no rematching, college c's expected payoff after observing signal x_c is the same under both information structures (say, u^*). Now suppose that in the coarser structure there is a strategy for c that gives it a strictly higher payoff, $u > u^*$. Then the same strategy is available to c under the finer information structure, and gives it the same payoff u, contradicting the assumption that u^* was the highest expected payoff c could obtain in that case. □

Proof of Theorem 5. For the proof, we can focus on the case where (1) v_{c_t} is strictly decreasing in t and (2) $v_{c_t} > 0$ for all t. The first assumption is an innocuous relabeling, while the second one can be made because colleges with $v_{c_t} < 0$ can be excluded from the analysis without altering anything. Also, we will prove a slightly stronger result: mechanism μ^{SD} is stable under information structure $\mathbf{z}$ that reveals to every college c_t its own match as well as the matches of all colleges $c_{t'}$ with $t' < t$. By Theorem 1, this result implies that μ^{SD} is weakly stable.

Consider the following profile of strategies and beliefs in game $\Gamma(\mu^{SD}, \mathbf{z})$:

1. Each college c_t reports its signal truthfully.
2. After observing its own match and the matches of colleges $c_{t'}$ for $t' < t$, college c_t updates its beliefs about the signals x and the draw of random variable ω using Bayes' rule, conditional on its own signal x_{c_t}, its own report $\hat{x}_{c_t}$, and on the assumption that all other colleges reported their signals truthfully. Note that, just like in the proof of Theorem 1, due to the full support assumption, this updating is always feasible, because every combination of colleges' signals is possible.

3. If college c_t reported its signal truthfully, it does nothing in the rematching stage. If it misreported its signal, it chooses the optimal strategy in the rematching stage (do nothing, drop a student, or make a rematching offer to one student), under the assumptions that (1) other colleges do nothing in the rematching stage and (2) any student not matched with college $c_{t'}$ for $t' < t$ will accept a rematching offer of college c_t.
4. Students accept a rematching offer if and only if it is better than their assigned match, and if they have multiple rematching offers that are better than their assigned match, they pick the most preferred one.

Let us show that this candidate profile of strategies and beliefs forms a perfect Bayesian equilibrium. By construction (part 2 above), this profile is consistent, i.e., the beliefs are updated correctly given the strategies. We need to show that it is also sequentially rational, i.e., the strategies are optimal given the beliefs.

Fix college c_t and note that the matches (both those generated by μ^{SD} and the final ones in game Γ) of colleges $c_{t'}$ for $t' < t$ do not depend on the strategy of college c_t, because (i) college c_t's report is not used by μ^{SD} in producing those matches and (ii) all of college c_t's rematching offers to the students matched to colleges $c_{t'}$ for $t' < t$ will be rejected by all those students in Step 6 of Γ.

We now show that college c_t would not be able to benefit by misreporting and/or rematching even if it knew in advance (prior to reporting its signal) the final matches of colleges $c_{t'}$ for $t' < t$. By the law of iterated expectations, this will imply that it would not be able to benefit by misreporting and/or rematching if it did not know those matches in advance.

Suppose college c_t knows that colleges $c_{t'}$ for $t' < t$ will receive matches $m(c_{t'})$. If it reports its signals truthfully and does not subsequently rematch, it will receive its most preferred subset (subject to its capacity constraint) from the remaining set of students, conditional on its own signal and on matches $m(c_{t'})$; in the case of ties, it will receive one of its most preferred subsets. Denote the expected payoff from this strategy as u^*. Consider any other strategy of college c_t (say, σ), and suppose it results in $J \geqslant 1$ possible matches for c_t, with a particular match m_j, $j = 1, \ldots, J$, occurring with probability p_j, $\sum_{j=1}^{J} p_j = 1$. (The number of possible matches for c_t is finite, because the number of students, and hence the subsets of students, is finite. The randomness can come from two sources: the randomization by the mediator in the case of ties and the use of a mixed strategy by college c_t.) Crucially, the outcome produced by this strategy for college c_t does not depend on the strategies or signals of colleges $c_{t''}$ for $t'' > t$, because (i) the reports of these colleges are not used by μ^{SD} to produce the match of c_t and (ii) all of their rematching offers to students matched to c_t will be rejected by all those students. Note also that by construction of μ^{SD}, the outcome does not depend on any information of colleges $c_{t'}$ for $t' < t$ beyond that revealed by their matches. Thus, the expected payoff of college c_t from using strategy σ is equal to

$$\sum_{j=1}^{J} p_j u_{c_t, m_j}\left(x_{c_t}, \{m_{c_{t'}}\}_{t'<t}\right),$$

which is less than or equal to u^* because $\sum_{j=1}^{J} p_j = 1$ and by the definition of u^*, for every j, $u_{c_t, m_j}(x_{c_t}, \{m_{c_{t'}}\}_{t'<t}) \leqslant u^*$. This shows that no college can benefit from employing an alternative strategy σ at any stage of Γ, establishing that the candidate profile of strategies and beliefs described above forms a perfect Bayesian equilibrium. □

References

[1] R. Aumann, Agreeing to disagree, Ann. Statist. 4 (1976) 1236–1239.
[2] H. Chade, Matching with noise and the acceptance curse, J. Econ. Theory 129 (2006) 81–113.
[3] H. Chade, G. Lewis, L. Smith, The college admissions problem with uncertainty, Mimeo, 2007.
[4] V. Crawford, Comparative statics in matching markets, J. Econ. Theory 54 (1991) 389–400.
[5] L. Dugatkin, Sexual selection and imitation: Females copy the mate choice of others, Amer. Naturalist 139 (1992) 1384–1389.
[6] F. Forges, Posterior efficiency, Games Econ. Behav. 6 (1994) 238–261.
[7] D. Gale, L. Shapley, College admissions and the stability of marriage, Amer. Math. Monthly 69 (1962) 9–15.
[8] B. Holmstrom, R. Myerson, Efficient and durable decision rules with incomplete information, Econometrica 51 (1983) 1799–1820.
[9] H. Hoppe, B. Moldovanu, A. Sela, The theory of assortative matching based on costly signals, Rev. Econ. Stud. 76 (2009) 253–281.
[10] S.-H. Lee, Early admission program: Does it hurt efficiency?, in: Three Essays on Applied Microeconomics, Ph.D. thesis, University of Pennsylvania, 2004, Ch. 1.
[11] J. Ma, Stable matchings and rematching-proof equilibria in a two-sided matching market, J. Econ. Theory 66 (1995) 352–369.
[12] R. Myerson, Incentive compatibility and the bargaining problem, Econometrica 47 (1979) 61–73.
[13] R. Myerson, Generalized principal-agent problems, J. Math. Econ. 10 (1982) 67–81.
[14] E. Nagypal, Optimal application behavior with incomplete information, Mimeo, 2004.
[15] A. Roth, Two-sided matching with incomplete information about others' preferences, Games Econ. Behav. 1 (1989) 191–209.
[16] A. Roth, M. Sotomayor, Two-Sided Matching, Cambridge University Press, Cambridge, MA, 1990.
[17] R. Wilson, Game-theoretic analyses of trading processes, in: T. Bewley (Ed.), Advances in Economic Theory: Fifth World Congress, Cambridge University Press, Cambridge, 1987.