Rev. Econ. Design (2015) 19:3–24
DOI 10.1007/s10058-014-0156-3

ORIGINAL PAPER

Group stability in matching with interdependent values

Archishman Chakraborty · Alessandro Citanna · Michael Ostrovsky

Received: 31 March 2013 / Accepted: 3 January 2014 / Published online: 18 January 2014

Abstract We study two-sided many-to-one matching markets with interdependent valuations and imperfect information held by one side of the market. The other side has common and known preferences over potential mates. In this setting, pairwise stability does not imply group stability: mechanisms that are stable with respect to deviations by pairs of agents may be vulnerable to deviations by groups. We formalize a notion of group stability and construct a "modified serial dictatorship" mechanism that implements group stable matchings. We further discuss the robustness of our notion of stability and examine efficiency properties of modified serial dictatorship.

Keywords Interdependent values · Group stability · Modified serial dictatorship

JEL Classification C78 · D82

1 Introduction

In many matching markets agents have incomplete information about potential mates. Consequently, an agent's preferences over mates may change if she obtains or infers the information held by other agents. For instance, admissions officers may be unsure about

A. Chakraborty · A. Citanna (✉)
Yeshiva University, New York, NY, USA
e-mail: citanna@yu.edu

A. Chakraborty
e-mail: archishman@yu.edu

M. Ostrovsky
Stanford University, Stanford, CA, USA
e-mail: ostrovsky@gsb.stanford.edu

the quality of applicants to a graduate program and they may update their estimates if they learn that the applicant has been admitted to another program. Employers may face uncertainty about the skills and ability of prospective hires and they may revise their opinions upon learning the information held by other employers.[1]

Under incomplete information and interdependent values, the information obtained during the operation of the market may lead an agent, or a coalition of agents, to revise their valuations and challenge the market outcome in favor of an alternative that is preferred by the coalition. Since agents anticipate this possibility and take it into account in their behavior, this places additional constraints on what a mechanism designer can do in terms of aggregating agents' information and deciding on allocations. It is therefore of interest to identify mechanisms under which matching markets with interdependent values are stable, i.e., immune to misrepresentation of information and to objections raised by agents after the operation of the market.

Chakraborty Citanna, and Ostrovsky (2010; hereafter CCO) introduce a notion of stability for two-sided problems of students and colleges where colleges have interdependent values over applicants. Stability in CCO mimics the traditional notions of stability in marriage markets: an assignment is stable if no individual agent (a college or a student) wants to unilaterally drop its assigned partner and no unmatched college–student pair could benefit from matching with each other (and possibly dropping their assigned partners). In the classical many-to-one matching setting with private values, this notion of *pairwise* stability is equivalent to a (potentially more restrictive) notion of *group* stability, under which larger coalitions are allowed to form and rematch (Roth and Sotomayor 1990, p. 130).[2] In this paper we show that under interdependent values, pairwise stability is not sufficient to guarantee group stability. A pairwise stable matching mechanism may be vulnerable to a group of colleges and students proposing an alternative assignment that they unanimously prefer after observing their matching outcomes.

CCO focus on matching markets with one-sided incomplete information, in which college qualities are known and student preferences over colleges are common knowledge. Furthermore, students have no private information about their quality and the student side of the matching market has homogeneous preferences.[3] In contrast, colleges have signals about the students. These signals could be informative about the student on different dimensions such as analytical ability, athletic talents, communication skills, or leadership potential. Colleges may have arbitrary, even opposed

[1] As an illustration, consider the following quote from a corporate recruiter about the market for young financial analysts: "If you're far along with Blackstone or K.K.R. and you're interviewing with Carlyle, they'll tell you to absolutely name-drop. Firms want to make sure they are picking a candidate who is loved by other firms as well." (New York Times Dealbook, http://dealbook.nytimes.com/2013/05/23/a-rush-to-recruit-young-analysts-only-months-on-the-job/, May 23, 2013.)

[2] This statement is true in the absence of complementarities, which is what we also assume throughout the paper.

[3] Such restrictions are necessary for positive existence results [see Roth (1989) and CCO]. They are also natural in many matching settings. For instance, student preferences over colleges and professional schools are often determined by rankings provided by newspapers and magazines. One-sided preference homogeneity is also an assumption frequently made in the directed search literature; see Shi (2002) for a list of references.

preferences over one or more of these dimensions. Crucially however, valuations are interdependent and a college's preferences over different students may depend on the signals held by other colleges.

We share with CCO these aspects of the environment but look for direct centralized matching mechanisms that cannot be blocked by any coalition once the matching outcome is determined. Such mechanisms must also be incentive compatible. More precisely, colleges cannot have an incentive to lie to the mechanism designer even after anticipating the possibility of subsequent rematching. We call these mechanisms *group stable*. We show that the set of group stable mechanisms is always nonempty (Theorem 1) by constructing a stable mechanism that we call *modified serial dictatorship*. Under modified serial dictatorship, colleges are ordered in a descending sequence according to the students' common preferences. Each college is then assigned students based on the college's evaluation of them, conditional on the college's own private signals and on the *signals of colleges above* (but not below) them in the student's rankings. With such a matching rule, no subset of college or students can infer enough information from the observed outcome to allow it to profit via a blocking coalition.

Modified serial dictatorship is different from the simple serial dictatorship mechanism introduced in CCO. Under the latter, each college is assigned students based on its own signals and the matching outcomes of the better colleges, but not those colleges' signals. For pairwise stability, either mechanism works. In fact there is a whole family of related mechanisms that result in pairwise stable matchings (see CCO, p. 97, footnote 8). For group stability, however, the choice of a particular serial dictatorship mechanism matters. While modified serial dictatorship produces a group stable outcome, simple serial dictatorship does not always do so. Simple serial dictatorship may not be group stable because it does not use all the information held by higher ranked colleges to determine the allocation of lower ranked colleges. In environments where colleges value students in different ways at least on some dimensions, this unused information may leave unrealized some gains from trade among coalitions of colleges. By sharing this unused information, these coalitions may profitably rematch among themselves. We confirm this intuition in Example 1 after defining what it means for a coalition to object to an allocation.

Even under one-sided private information and homogenous student preferences, extending the analysis of stability from pairs to groups is not immediate. A suitable definition of group stability has to address three fundamental issues. The first, common also to the CCO setup, is *observability*, or the extent to which the status quo outcome is publicly observed at the posterior stage when coalitions may raise objections. The second is *coalition formation*, that is, what coalitions can be formed and what rematching protocols can be used. The third is *communication*, i.e., the restrictions, if any, that one imposes on information sharing among coalition members at the time they raise objections.

With respect to the observability issue, we follow CCO and consider weak observability—each agent only observes its own mates and does not directly observe the mates of others at the time they may raise objections. Observing the entire matching outcome (strong observability) leads to impossibility results demonstrated in CCO and therefore it is too demanding a requirement. With respect to coalition formation, we also follow CCO and assume that any coalition can be formed if agents find it in their interest to do so, i.e., there are no re-matching frictions.

Unlike CCO however, we allow coalitions larger than a college–student pair. In doing so we also need to specify the mechanisms available to blocking coalitions at the time they raise objections. We focus on a simple mechanism with the following features. First, we allow coalitions to object to the decision made by the status quo mechanism only via proposing an alternative matching (as opposed to a matching rule). Second, we model this counterproposal stage as a simultaneous voting game in which coalition partners must unanimously accept the alternative match for the objection to succeed and look at Bayesian Nash equilibria of this game. Third, we assume that agents can share private information only if they "belong to the coalition" and are made strictly better off by the counterproposal. Since agents will condition their acceptance of a counterproposal on their own private information (as well as on the information they have obtained from the initial allocation), a blocking coalition in general reveals to its members at least some of the information held by other members.

The existence of a group stable mechanism depends upon the restrictions we impose on blocking coalitions. In principle, a group of agents could design a general communication mechanism to share the information each agent in the group has in an incentive compatible manner. Our restriction that a blocking coalition can only engage in unanimous voting over an alternative deterministic match, as opposed to using a general mechanism, captures situations where a general mechanism may be too costly to operate for a blocking coalition at the stage when the status quo rule has already proposed a match. These costs may be direct or take the form of constraints in designing complex mechanisms that require precise specification of randomization schemes and message games (for a similar point see Forges 1994). Whether or not these restrictions are realistic, we show that such restrictions are in fact necessary to obtain existence of group stable mechanisms (Theorem 2). Necessity gives our definitions and results a normative content and provides a tight characterization of group stability.

The principal force behind this tight necessity result is the possibility of anticipated renegotiation. Anticipated renegotiation is a key consideration shared with CCO that goes beyond the twin requirements of incentive compatibility on one hand and immunity to objections given truthful information revelation on the other. Under anticipated renegotiation, agents are allowed not only to lie to the mechanism or block the observed matching outcome, but they can also lie to the mechanism anticipating a subsequent objection raised after observing the matching outcome. Immunity to anticipated renegotiation is a natural requirement that stable matching rules should satisfy in situations where agents are not only aware that the mechanism needs to aggregate information but are also aware that they can subsequently object to the proposed outcome.

We conclude the paper by considering the efficiency properties of modified serial dictatorship. We show that in environments where higher-ranked colleges have higher thresholds for admitting students, modified serial dictatorship always produces an ex-post Pareto efficient matching. In this it has better efficiency properties than the simple serial dictatorship proposed by CCO.

The literature on matching has usually considered private value environments (see, e.g., Roth and Sotomayor 1990, for a survey of many classical results). In a different context of positive analyses of specific matching markets with frictions, Chade (2006), Hoppe et al. (2009), and Ely and Siegel (2013) also highlight the role of information revealed by observed matchings on the operation of the market. We employ instead a mechanism design approach which bears similarities to the notion of truthful and obedient behavior in Myerson (1982). In the mechanism design vein, Neeman and Pavlov (2013) have recently looked at renegotiation-proofness in allocation problems with transfers. As mentioned above, CCO consider pairwise stability only, and this paper should be seen as an extension of the analysis carried out there. In particular, a key innovation arising in both papers is the notion of anticipated renegotiation, which has subsequently been applied in private-value settings by Kojima (2011) and Afacan (2012). This paper also relates to the literature on cooperative games, specifically, on the core under incomplete information (e.g., Wilson (1978), Vohra (1999), Forges et al. (2001); for a survey, see Forges et al. (2001)). Our analysis differs because we focus on objections at the posterior stage, and we consider the possibility of anticipated renegotiation.[4]

2 The matching market

We study two-sided, many-to-one matching problems, or "matching markets," where students can attend at most one college at a time, while colleges can possibly admit multiple students. We follow CCO in defining such matching markets.

The agents are students and colleges. The set of students is denoted by $\mathbf{S} = \{1, \ldots, S\}$, with $S \geq 1$ and typical element s, whereas the set of colleges is denoted by $\mathbf{C} = \{1, \ldots, C\}$, with $C > 1$ and typical element c. Each college has a capacity of $k_c > 0$ students.

Students differ in their unobserved quality or ability $q_s \in \mathbf{Q}$, a finite set. Neither colleges nor students know the quality q_s of any student s, but each college c receives a private signal $x_{c,s} \in \mathbf{X}_c$, a finite set. We may think of $x_{c,s}$ as the outcome of a privately observed informative test or interview for student s . Let $x_s = (x_{c,s})_{c\in\mathbf{C}}$ be the vector of signals associated with each student s, and $x_c = (x_{c,s})_{s\in\mathbf{S}}$ be the vector of signals received by each college c, so that $x = (x_s)_{s\in\mathbf{S}} = (x_c)_{c\in\mathbf{C}}$ represents the private information available overall and $x \in \mathbf{X} = \times_c \mathbf{X}_c^S$. We let Pr be the joint probability distribution over signals and qualities $\mathbf{X} \times \mathbf{Q}^S$, and assume that for any $x, q, \Pr(x, q) > 0$ (full support).

Students have preferences over single mates and their preferences are state-independent and common knowledge. Also, they strictly prefer to be matched to any college rather than staying unmatched. Formally, letting $v_{s,k}(x)$ be a student utility from matching with agent k at signals x , we assume that for each s,

A1.1 $v_{s,k}(x) = v_{s,k}$ for all k; $v_{s,c} = v_c$, with $v_c > v_{c'}$ if $c > c'$; and $v_c > v_{s,s} = 0$ for all c.

Colleges have preferences over groups of mates which are additively separable, and for each c,

[4] Forges (1994) also studies the posterior stage, but only for efficiency and without considering anticipated renegotiation.

A1.2 $w_{c,s}(x, q)$ is the payoff to college c from accepting student s when qualities are q and signals are x, and $w_{c,c} = 0$. We let $u_{c,s}(x) = \sum_q w_{c,s}(x, q)\Pr(q|x)$.

Note that college c's payoff from student s does not depend on who else is admitted to that college, a special case of responsiveness (see Roth and Sotomayor 1990, Definition 5.2), and that agents have expected utility preferences.

We define matchings not only for all colleges and students, but also for coalitions made of subsets of agents. Formally, let $\mathcal{C} = (\mathbf{C}', \mathbf{S}')$ be a coalition with $\mathbf{C}' \subset \mathbf{C}$ and $\mathbf{S}' \subset \mathbf{S}$, and $\mathbf{C}' \cup \mathbf{S}' \neq \emptyset$. A *(two-sided many-to-one) matching m feasible for the coalition* $\mathcal{C}$ 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) for any student $s \in \mathbf{S}'$, $m(s) = c$ 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 $\mathbf{M}_{\mathcal{C}}$ be the set of all matches feasible for $\mathcal{C}$. When $\mathcal{C}$ is the grand coalition $(\mathbf{C}, \mathbf{S})$, we simply write $\mathbf{M}$ for the set of all feasible matchings.

Finally, let $\mathbf{F}_c$ be information college c has, a coarsening of $\mathbf{X}$, such that $\Pr(\mathbf{F}_c) > 0$. For any agent k and matches $m(c)$ and $m'(c)$ we define

$$u_{c,k}(\mathbf{F}_c) = \sum_x u_{c,k}(x) \Pr(x|\mathbf{F}_c);$$

$$U_c(m(c)|\mathbf{F}_c) = \sum_{s \in m(c)} u_{c,s}(\mathbf{F}_c); \text{ and}$$

$$U_c(m(c) - m'(c)|\mathbf{F}_c) = U_c(m(c)|\mathbf{F}_c) - U_c(m'(c)|\mathbf{F}_c).$$

3 Definition of group stability

Since agents have private information, this information may have to be aggregated to arrive at a matching outcome for the market described above. We take a mechanism design approach where agents report their signals to a mediator who subsequently proposes a matching for the market. Formally, a *direct revelation matching mechanism* μ is a function from the set $\hat{\mathbf{X}} \times [0, 1]$ of reported signal profiles $\hat{x}$ and draws of a random variable $\omega \in [0, 1]$ to the set $\mathbf{M}$ of matchings. The presence of ω allows the final matching to be stochastic, i.e., the mediator can randomize. Without loss of generality, we assume that ω is distributed uniformly on $[0, 1]$. When $\mu(\hat{x}, \omega)$ does not depend on ω for all $\hat{x}$, we say that μ is *deterministic*; while if $\mu(\hat{x}, \omega)$ does not depend on $\hat{x}$ we say that it is *constant*.

We are interested in the possibility of objections raised by coalitions of agents to the matching outcome proposed by μ. These objections are raised at the posterior stage after some information about the outcome proposed by μ has been observed by the agents. Because of interdependent values, agents may use the information revealed by a matching mechanism to object to its outcome. Since the mechanism itself affects what is known at the posterior stage when objections can be raised to the mechanism, we need to define stability on matching mechanisms, not simply on matchings (see CCO for a fuller discussion). We also focus attention on the case of weak observability where each college only observes the portion $m(c)$ of the matching outcome that pertains to that college, and similarly each student only observes the college $m(s)$ that it is matched to. As shown in CCO, even pairwise stable matching mechanisms (let alone group stable ones) do not exist in general when agents observe more information about the proposed matching at the posterior stage when they can raise objections.

We now turn to the key definition of what we mean by an objection by a coalition to the outcome of the status quo mechanism μ. A coalition $\mathcal{C}$ objects to match m^* realized after messages $\widehat{x}^*$ have been reported to the mechanism and the signal realization is x^*, by proposing some alternative match m'. Given such a coalition and an alternative, coalition members engage in unanimous voting on the alternative. The following definition makes precise our idea of an objection or block. For given matchings m and m', with $m' \in \mathbf{M}_{\mathcal{C}}$, and any $c \in \mathbf{C}$ we let

$$\begin{aligned}\mathbf{M}_c(m, m') = \{m''(c) :\ & \left|m''(c)\right| \leq k_c, m'(c) \cap S \subset m''(c) \\ & \text{and if } s \in m''(c) \backslash (m'(c) \cap S) \text{ then } s \in m(c)\}\end{aligned}$$

as being the available matches to college c relative to the pair m, m'. An element $m''(c)$ of $\mathbf{M}_c(m, m')$ contains all the students in $m'(c)$ with any remaining slots left empty or filled with students in $m(c)$. Let $\mathbf{X}_k$ be agent k's set of signals, and $\hat{\mathbf{X}}_k$ the agent's set of messages (with $\hat{\mathbf{X}}_k = \mathbf{X}_k \equiv \{0\}$ if $k \in \mathbf{S}$).

Definition 1 A mechanism μ can be "blocked in principle" by a coalition $\mathcal{C}$ at $(\widehat{x}^*, x^*, m^*)$ if there exist an alternative match m' and voting rules $\alpha_k : \hat{\mathbf{X}}_k \times \mathbf{X}_k \times \mathbf{M}_k \rightarrow \{0, 1\}$ for each $k \in \mathcal{C}$, satisfying

$$\begin{aligned}&\alpha_c(\widehat{x}_c, x_c, m(c)) = 1 \\ &\quad iff \max_{m''(c) \in \mathbf{M}_c(m, m')} U_c(m''(c) - m(c) | \widehat{x}_c, x_c, m(c), \alpha_{-c} = 1, \widehat{x}_{-c} = x_{-c}) > 0\end{aligned} \tag{PS1}$$

for each $\widehat{x}_c, x_c, m(c)$ and each $c \in \mathbf{C}'$; and

$$\alpha_s(\widehat{x}_s, x_s, m(s)) = 1 \quad \text{iff} v_{s,m'(s)} > v_{s,m(s)} \tag{PS2}$$

for each $\widehat{x}_s, x_s, m(s)$ and each $s \in \mathbf{S}'$, with $\alpha_k(\widehat{x}_k^*, x_k^*, m^*(k)) = 1$ for all $k \in \mathcal{C}$.

Definition 1 evaluates a block for an agent assuming the agent is pivotal, i.e., all other agents in the coalition have accepted the alternative ($\alpha_{-c} = 1$) and assuming also that other agents in the coalition reported truthfully their signals to the mediator ($\widehat{x}_{-c} = x_{-c}$).[5] The restriction to $\mathbf{M}_c(m, m')$ in (PS1) is just a feasibility requirement relative to the coalition a college belongs to. For instance, if a college blocks alone (i.e., $\mathbf{S}' = \varnothing$), it can only drop some students it was assigned to under μ.

[5] We use the subscript $-c$ to denote all members of the coalition $\mathcal{C}$ apart from c. When evaluating a block, the agent also assumes that all other agents not in the coalition $\mathcal{C}$ have reported their signals truthfully, but we suppress this extra notation.

Notice from Definition 1 that all members of the coalition must strictly prefer to accept the blocking allocation m', and that m' is a deterministic and constant mechanism $\mu'(\hat{x}, \omega) = m'$, all $\hat{x}, \omega$. This is identical to requiring that each member of a blocking coalition strictly prefers to the status quo a feasible and non-random alternative matching *outcome*, as opposed to a matching *rule*. These are restrictions on the ability of blocking coalitions to "effectively negotiate" (Myerson 1991, Ch. 9–10). Especially when observing m does not fully reveal the state of the world x, one could imagine that a blocking coalition may be able to design a communication game and thereby implement a general incentive compatible alternative mechanism μ', neither constant nor deterministic (see, e.g., Definition 3.2 in Forges 1994). In contrast, Definition 1 allows communication only to the extent of what can be inferred from the equilibrium of the acceptance game, i.e., from conditioning on the event that agent k is pivotal given strategies $\alpha_{\mathcal{C}}$.[6]

Let $U_k(\mu; \widehat{x}_k, x_k)$ be agent k's interim expected payoff when agent k reports $\widehat{x}_k$, his signal is x_k, all other agents report signals truthfully and the matching assigned by μ is accepted. In addition, let $U_k(\mu'; \widehat{x}_k, x_k, \alpha_{\mathcal{C}})$ be agent k's interim expected payoff when his report is $\widehat{x}_k$, his type x_k and all coalition members in $\mathcal{C}$ vote on μ' according to $\alpha_{\mathcal{C}}$ (assuming all agents other than k report their signals truthfully). We are now ready to define group stability.

Definition 2 A mechanism μ is group stable if it satisfies the following three conditions:

1. μ is incentive compatible: $x_k \in \arg\max_{\widehat{x}_k} U_k(\mu; \widehat{x}_k, x_k)$ for each x_k and each k;
2. μ cannot be blocked in principle by any coalition $\mathcal{C}$ at any $(\widehat{x}, x, m)$ with $\widehat{x}_k = x_k$ for all k;
3. if μ is blocked in principle by some coalition $\mathcal{C}$ at $(\widehat{x}, x, m)$ satisfying $\widehat{x}_k \neq x_k$ for a unique $k \in \mathcal{C}$, then $U_k(\mu'; \widehat{x}_k, x_k, \alpha_{\mathcal{C}}) \leq U_k(\mu; x_k, x_k)$.

Condition 1 in Definition 2 is the usual notion of incentive compatibility, i.e., no agent has an incentive to misreport his signal assuming no other agent does so, assuming also that the matching proposed by the mechanism will be the final outcome. Condition 2 states that no coalition can object on the path of play, i.e., given all agents have reported their signals truthfully. Condition 3 goes beyond these two restrictions by taking into account the possibility of anticipated renegotiation, namely, that an agent may lie to the mediator and then block at the posterior stage. Overall, Definition 2 makes sure that "truth telling and no rematching" dominates not only "truth telling and rematching" (as guaranteed by Condition 2) and "lying and no rematching" (as guaranteed by Condition 1), but also "lying and rematching" (as guaranteed by Condition 3). The classical literature on stable matchings that are also incentive compatible does not consider the requirement of no anticipated renegotiation, a key innovation

[6] If we defined objections under strong observability, under which every agent observes the entire matching at the time of raising objections, Definition 1 would have to be modified only in letting strategies α_c and α_s to be functions of the entire match m, as opposed to of only $m(c)$ and $m(s)$, respectively. When we do so, we refer to the mechanisms as being *strongly stable*.

shared by this paper and CCO.[7] We show below that it will play a crucial role in some of the results below—in particular, in showing why the restricted communication requirements embedded in Definition 1 are necessary for the existence of group stable mechanisms.

CCO define pairwise stability in terms of the perfect Bayesian equilibria (PBE) of a two-stage non-cooperative game of signal reporting and rematching. In this game, colleges first report their signals to a mediator who outputs a proposed matching according to a mechanism μ. Colleges may then simultaneously make offers to students they are not matched with (or simply drop a student they are currently matched with). In contrast, Definition 2 bypasses the full description of a two-stage game of signal reporting and rematching and directly provides conditions for a mechanism to be stable. While this approach is at first sight quite different from the one in CCO, the two definitions are equivalent when we restrict attention to pairwise stability, i.e., in the case where a coalition can consist of a single college and student or a subset thereof.

To see this, consider first a situation where a matching mechanism μ is not a PBE of the two stage game of reporting and rematching proposed by CCO and so not pairwise stable in their sense. Suppose, in particular, that μ is not a PBE because in the second stage of the game, after all realized signals x^* are revealed truthfully and a matching outcome m^* is obtained, some college c profitably deviates by making an offer to some student s that c is currently not matched with. College c must prefer this alternative knowing its own signal x_c^*, its own truthful report $\widehat{x}_c^* = x_c^*$, and its own match $m^*(c)$. In addition, college c should evaluate this profitable deviation also by conditioning on the fact that the student will accept its offer only when college c is preferred by s to his current match, assuming also that all other colleges have reported their signals truthfully, $\widehat{x}_{-c}^* = x_{-c}^*$. But then the mechanism μ is not group stable in the sense of the this paper because it violates condition (2) in Definition 2. The relevant coalition is $\mathcal{C} = \{c, s\}$ and the alternative μ' is that s match with c. This block will succeed in the state of the world $(\widehat{x}^*, x^*, m^*)$ with $\widehat{x}_k^* = x_k^*$ for all k. Condition (PS2) in Definition 1 states that the student s prefers to accept the alternative only when c is preferred to its current match. Condition (PS1) states that college c prefers the alternative given the information contained in its own signal and observed match, conditional on the student's expected behavior and assuming truthful reporting by all parties including c. These conditions are identical to those underlying the profitable deviation by college c in the two-stage game used by CCO.

Similarly, if μ is not a PBE because a college simply prefers to misreport in the first stage of the game proposed in CCO, condition (1) in Definition 2 will be violated. On the other hand, if a college prefers to misreport in the first stage, anticipating subsequent rematching in the second stage, condition (3) in Definition 2 will be violated, with the relevant coalitions consisting of college c and possibly one or more students. In the other direction, if mechanism μ violates any of conditions (1)–(3) in Definition 2, with blocks restricted to be singletons or college-student pairs, then there is a corresponding profitable deviation for some college in the two-stage game of reporting and rematching used in CCO.

[7] See Kojima (2011) and Afacan (2012) for applications of the no anticipated renegotiation requirement to private value settings.

In our consideration of group stability, we avoid characterizing stability in terms of the PBE of a non-cooperative game not only for reasons of simplicity but also to avoid obtaining group stability trivially. With coalitions of multiple agents and unanimous voting on an alternative, a mechanism μ may be a PBE of a (suitably specified) two stage game of signal reporting and rematching simply because no member of an objecting coalition votes in favor of the alternative expecting no other member to do so. Definition 1 eliminates the possibility of such coordination failures between coalition members.[8] It imposes the stronger requirement that μ is stable only if *there is no* profile α_C that a coalition can use to block.

With the definition of group stability in place, we turn to our first example. In the example we apply the serial dictatorship mechanism provided by CCO to characterize a pairwise stable matching mechanism. We show that this matching mechanism may not be group stable. Next, we propose a modification of the CCO mechanism that will generate group stable allocations in the example. Subsequently, in Sect. 4 we show that this modification delivers a group stable matching mechanism in all environments.

Example 1 There are two colleges 1 and 2 and one student s. The student prefers college 2 to college 1 and prefers matching with any college to remaining unmatched, i.e., $v_2 > v_1 > v_s$. Each college $c = 1, 2$ receives an informative binary signal $x_c = \{L, H\}$, each equally likely. We suppose that (i) college 1 prefers to match if and only if both $x_1 = H$ and $x_2 = H$, and based only on its own signal college 1 prefers to match with the student if and only if $x_1 = H$; (ii) college 2 prefers to match with the student if and only if $x_1 = H$ and $x_2 = L$, and based only on its own signal college 2 prefers staying unmatched over matching with the student.

In Example 1, college 1's signal may measure the student's reasoning ability while college 2's signal may measure the student's inclination to take risks. Both colleges have similar preferences over the student's reasoning ability but have opposing preferences on the student's risk attitudes—while college 1 values a risk taker, college 2 does not. The opposing preferences on the second dimension is what leads the pairwise stable allocation rule obtained via simple serial dictatorship not to be group stable.

Simple serial dictatorship works as follows. First, the student is evaluated by the most preferred college 2 given its own signal. Since college 2 prefers not to admit the student given only its own signal, the student is next evaluated by college 1. College 2's decision to not admit the student is uninformative about its signal and so college 1 learns no information from the fact that the student is available. Since college 1 is willing to admit the student only when $x_1 = H$, the student is admitted to college 1 in this case. When $x_1 = L$, the student remains unmatched. This matching rule has the property that no college has an incentive to lie about the signal received, whether or not it subsequently tries to rematch, and indeed no college will try to rematch. Simple serial dictatorship is pairwise stable.

[8] With pairwise stability and known student preferences this possibility cannot arise—if some preferred college makes an offer to a student, the student will always accept.

But this matching rule is not group stable. To see this, consider the grand coalition of the two colleges and the student that block using the alternative proposal that the student match with college 2. The student is always willing to accept such an alternative regardless of what he gets under the status quo mechanism since college 2 is the student's most preferred college. Suppose college 2 accepts this block only when $x_2 = L$ and college 1 only when $x_1 = H$. Since college 2 prefers to accept the student only in the state of the world $(x_1, x_2) = (H, L)$, college 2 is strictly better off from accepting the block if and only if $x_2 = L$, conditional on being pivotal and using the information contained in college 1's acceptance strategy. Similarly, since college 1 prefers to accept the student when $(x_1, x_2) = (H, H)$ and prefers staying unmatched otherwise, college 1 is also strictly better off from accepting the block if and only if $x_1 = H$, conditional on being pivotal and using the information contained in college 2's acceptance strategy. In essence, the block succeeds in the state of the world $(x_1, x_2) = (H, L)$ because both colleges and the student prefer that the student be matched with college 2 in that state. By conditioning their acceptance decisions on their private signal and on being pivotal, the two colleges can in effect communicate their private information to the other party. This shows that the serial dictatorship algorithm proposed by CCO does not in general rule out the possibility of information sharing among coalitions involving multiple colleges.

In the next section, we show that a modification of the serial dictatorship algorithm of CCO delivers group stable matching rules in every environment. In the present example, this modified serial dictatorship algorithm works as follows. As with simple serial dictatorship, the student is first evaluated by his most preferred college 2 given its own signal. Since college 2 prefers not to admit the student given only its own signal, the student is next evaluated by college 1. But unlike simple serial dictatorship, college 1 is asked to evaluate the student given not only its own signal but also the signal of college 2. This generates the following matching mechanism: the student is matched with college 1 when $(x_1, x_2) = (H, H)$ and remains unmatched otherwise. As Theorem 1 shows, this assignment rule is group stable. Since modified serial dictatorship uses more information than simple serial dictatorship in generating matches, it eliminates the possibility of successful objections and information sharing by any arbitrary coalition.[9]

4 Main results

Our main results are twofold. First, we establish the existence of group stable mechanisms. The proof is constructive and uses the *modified serial dictatorship* algorithm outlined in Example 1 to deliver a mechanism with such properties. Second, we show that our definition of group stability is *tight*—group stable mechanisms fail to exist in general if we allow coalitions greater flexibility in sharing information and splitting the gains from participating in blocks. In this sense, we identify necessary and sufficient conditions for the existence of group stable mechanisms.

[9] Notice that modified serial dictatorship does not deliver an ex-post efficient allocations in Example 1. We discuss the efficiency properties of modified serial dictatorship in Sect. 5, comparing it also with simple serial dictatorship.

4.1 Existence of group stable mechanisms

Modified serial dictatorship works as follows. For each student s, the student's value to each college c is gauged using college c's own signal and *the signals possessed by all colleges above c in the students' (common) preference ordering*. Then, students are ranked according to their values. Let $\hat{k}_c$ be the number of students who carry positive value and are still available, i.e., were not assigned to any of the higher colleges in the student's rankings. *Strictly* order the students from best to worst, breaking ties among equal value students via uniform randomization. If $\hat{k}_c < k_c$, the mechanism assigns all the $\hat{k}_c$ students to college c leaving the remaining slots unassigned. Otherwise, the top k_c of the $\hat{k}_c$ students are assigned to college c. Then, the next college is considered, and the procedure repeated, starting from the top college and moving down the students' rankings. We denote the resulting mechanism as μ^{MSD}.

Theorem 1 *Mechanism μ^{MSD} is group stable.*

All proofs are in the Appendix. The main hurdle in the proof is to rule out group rematches. In principle, colleges that are higher in a student's ranking may want to obtain information from lower-ranked colleges by attempting to swap students with those colleges, a group rematching strategy. However, under A1.1 the lowest college in a student's ranking is lowest in all students' rankings. So the only way such a swap can work is when the lowest ranked college in a coalition obtains a valuable student who was matched with an even lower ranked college outside the coalition (or was unmatched); or if the lowest ranked college in a coalition learns information from the other colleges that allows it to profitably discard a previously accepted student. But since μ^{MSD} uses all the information available to higher ranked colleges to generate the matchings for lower ranked colleges, the lowest-ranked college in any coalition cannot obtain any information from other higher ranked colleges that it did not already have. This establishes Condition 2 of stability for μ^{MSD}. Indeed μ^{MSD} cannot be blocked if, at the blocking stage, in addition to the match $m(c)$ realized at x via μ^{MSD}, each college is also allowed to observe the actual signals of all colleges above it in a student's ranking. Furthermore, μ^{MSD} is incentive compatible for each college c (Condition 1), not only in interim terms, but also conditional on the actual signals received by all colleges $c' > c$, given truth-telling by colleges $c' < c$. Thus, it is ex-post incentive compatible for college 1 and, given truth-telling by college 1, incentive compatible for college 2 even given the actual signals of all colleges above 2, and so on. From here, full support of Pr immediately implies that μ^{MSD} satisfies Condition 3.[10]

[10] Modified serial dictatorship delivers group stability because, by construction, no college is able to infer anything from its own match about the information held by colleges below it in the students' ranking. Instead, under strong observability, by observing the entire match a college may infer valuable information about the signal obtained by colleges below it in the student's ranking, and so profitably block μ^{MSD}. As shown in CCO (Example 2) this ability to deduce the information held by lower colleges in the context of strong observability leads to the nonexistence of any (even only interim) incentive compatible strongly stable mechanism. Exactly the same intuition underlies the nonexistence of incentive compatible ex–post stable mechanisms.

4.2 The necessity of restrictions on blocks

Our definition of stability imposes a number of restrictions on the ability of blocking coalitions to raise objections. For ease of reference we list them as follows:

R1 No member of a blocking coalition can only be weakly better off from participating in a block.
R2 Blocking coalitions cannot propose a stochastic alternative μ' that is a probability distribution over different matches.
R3 Blocking coalitions cannot propose mechanisms that specify alternative matches as a function of messages exchanged by coalition members.

In this section we prove that these restrictions are all necessary for the existence of stable mechanisms in all matching markets. For instance, if we allow coalition members only to be weakly better off from participating in the block, then coalition members can "sell information at price zero" to other members, precluding the existence of stable mechanisms. Similarly, if a blocking coalition can propose lotteries over multiple feasible alternative matches, they can essentially create stochastic reward schemes and successfully block every possible mechanism. In the same spirit, if coalition members are allowed to use general mechanisms, then they can create randomization schemes by making matchings depend on payoff-irrelevant messages. This rules out the existence of stable rules when coalitions can use general mechanisms that are subject to self-selection and incentive compatibility constraints.

Theorem 2 *Group stable mechanisms do not exist in all matching markets if any one of the restrictions R1, R2 or R3 are relaxed.*

The proof of Theorem 2 consists of constructing matching markets where no group stable mechanism exists if either R1 or R2 or R3 is relaxed. Example 2 below is an example of such a matching market. In the proof we characterize the necessary properties of group stable mechanisms using, in particular the no anticipated renegotiation condition (Condition 3) of Definition 2, and show that no stable rule can exist unless one imposes all three restrictions on blocking coalitions listed above. To illustrate the intuition, we present here only the argument why μ^{MSD} fails to be stable in this market if any one of R1, R2 or R3 is relaxed.

Example 2 There are three colleges and one student s. The student prefers college 3 to college 2 to college 1 and prefers matching with any college to remaining unmatched, i.e., $v_3 > v_2 > v_1 > v_s$. Colleges 1 and 3 do not receive any informative signals. However, college 2 receives a binary informative signal $x_2 = \{L, H\}$, each equally likely. We suppose that (i) college 1 prefers to match with the student regardless of x_2; (ii) college 2 and college 3 prefer to match with the student if and only if $x_2 = H$; while (iii) in the absence of any information about college 2's signal, college 3 prefers to remain unmatched.

In this matching market, μ^{MSD} assigns the student to college 2 when $x_2 = H$ and to college 1 when $x_2 = L$. We begin by showing that μ^{MSD} can be blocked if we relax R1, allowing participation in blocks even if an agent is only weakly better off from doing so. More precisely, we replace the strict inequalities in (PS1) and (PS2) in Definition 1 for all agents in $\mathcal{C}$ with weak inequalities for all but at least one agent who should strictly prefer to accept. Definition 2 of a stable matching mechanism is accordingly modified and we use the term *group* stability* to denote this modified notion.

Modified serial dictatorship can be blocked if we use the notion of group* stability, via the coalition of the student and colleges 1 and 3 that uses the alternative that the student be matched with college 3. Since college 3 is the student's most preferred college, the student will always strictly prefer such an alternative. College 1 strictly prefers to reject such an alternative when the status quo μ^{MSD} proposes that the student be matched with college 1, i.e., when $x_2 = L$. However, college 1 weakly prefers to accept the alternative (in fact is indifferent) when college 1 observes that μ^{MSD} did not match the student with college 1. In such a case, college 1 can infer that μ^{MSD} has matched the student with college 2, so that $x_2 = H$. Since under group* stability we allow all but one agent to participate in a block even if they are only weakly better off from doing so, we let college 1 accept the alternative whenever μ^{MSD} has not matched it with the student. This acceptance strategy on the part of college 1 allows college 3 also to infer that, conditional on participation from college 1 , $x_2 = H$. Since college 3 strictly prefers to match with the student in such states of the world, it is strictly better off from participating in the block when $x_2 = H$. By allowing college 1 to participate in the block even if it is only weakly better off from doing so, we effectively allow it to convey to college 3 the information it obtains from observing its own part of μ^{MSD}. In contrast, under our original definition of a stable matching, college 1 would not be able to participate in this block since it does not strictly benefit from doing so.

A similar argument goes through if instead of R1 we relax R2 above. When alternatives are allowed to be lotteries, coalition members evaluate the alternative in expectation over possible lottery outcomes, and we call the resulting notion of stability *group** stability*. Under *group** stability* every coalition member must be strictly better off in expectation over the lottery outcomes proposed by μ' after taking into account the participation decisions of others.

We show now that *group** stability* allows coalition members to strictly benefit from blocking μ^{MSD} by considering again the coalition of the student and colleges 1 and 3. However, now we allow the alternative μ' to be the lottery that matches the student with college 1 with probability $p \in (0, 1)$ and with college 3 with probability $1 - p$, where p satisfies $pv_1 + (1 - p)v_3 > v_2$. Because of the restriction on p the student strictly prefers to accept the alternative when matched by μ^{MSD} to college 2 (or college 1). Since $p < 1$, college 1 strictly prefers the status quo when $x_2 = L$ and μ^{MSD} matches the student with college 1. But since $p > 0$, college 1 strictly prefers the alternative μ' when $x_2 = H$ and μ^{MSD} matches the student with college 2. Therefore, college 1 strictly prefers to accept the alternative if and only if it has not been matched with the student by μ^{MSD}. This allows college 3 to infer that $x_2 = H$ conditional on participation by college 1. Since college 3 strictly prefers to match with the student when $x_2 = H$, it is also strictly better off from accepting the alternative.[11]

[11] In the proof of Theorem 2, we only need to add to this argument by showing that a mechanism that never allocates the student to college 2 is vulnerable to anticipated renegotiation by that college. For instance, if

It is now easy to illustrate why the modification of blocks induced by a relaxation of R3 will also yield negative conclusions. To demonstrate this, we "replicate" Example 2 as follows. Suppose that there are two students, s and s', and three colleges 1, 2 and 3. The preferences and information of all agents are identical to Example 2. Moreover, college 2's signals for students are drawn independently across students. Also, each college has enough capacity to accept both students. With this replication, μ^{MSD} allocates any student to college 2 if that college's report for that student is H and allocates that student to college 1 otherwise. Coalitions are now allowed to propose deterministic alternative matches, but in a way that depends on messages sent by coalition members. It is then possible for a coalition of colleges 1 and 3 and the two students s and s' to each be strictly better off from objecting to μ^{MSD}. This can be achieved by designing an alternative proposal (a mechanism) that allocates student s (resp., s') to either college 1 or college 3 depending on a message sent by the other student s' (resp., s). Since each student's own match depends only on the message sent by the other student, each student is willing to randomize its messages in a manner that replicates the lottery used for establishing the instability of μ^{MSD} when R2 is relaxed. The arguments used above then extend to show that μ^{MSD} can be blocked via such a mechanism when college 2's signal for each student is H. Indeed, one can show that no stable mechanism exists when one allows coalitions to use general mechanisms in which the alternative allocation is determined as a function of messages sent by coalition partners.[12]

5 Efficiency properties of modified serial dictatorship

Modified serial dictatorship gives rise to matchings that are stable against the grand coalition at the posterior stage. If efficiency is defined relative to the information agents have at the posterior stage, then modified serial dictatorship is efficient provided dominance must be strict for every agent and the alternative is still restricted to be constant and deterministic. Similarly, if dominance only requires that an agent who shares information within the grand coalition has to strictly profit from the alternative matching, while other agents can be weakly better off, then again modified serial dictatorship is efficient. The reason is simple: μ^{MSD} cannot be blocked in principle by any coalition, whether agents outside the given coalition are better off or not. Hence, a fortiori it cannot be blocked in principle if we restrict such blocks not to hurt anyone outside the given coalition. Clearly, this rules out dominance as just defined. Example 2 has already shown that matchings obtained through modified serial dictatorship are

Footnote 11 continued
the mechanism allocates the student to college 3 when $x_2 = H$ and to college 1 when $x_2 = L$, then when $x_2 = H$ college 2 can profitably invert its signal and report $\widehat{x}_2 = L$ and then block in a coalition with the student.

[12] When R3 is relaxed, the alternative μ' is a mechanism that must not only satisfy self-selection constraints along the lines of conditions (PS1) and (PS2) in Definition 1, but also incentive compatibility constraints. These constraints are a straightforward, but notationally cumbersome, generalization to the posterior stage of Dutta and Vohra (2005) notion of the credible core. Therefore, we omit their formal development in the interest of brevity.

not guaranteed to be efficient if agents share information for free, that is, are allowed to participate in a coalition that makes them only weakly better off.

In light of Example 1, we also want to measure modified serial dictatorship against efficiency ex-post: μ is *ex-post dominated* at $x \in \mathbf{X}$ if there exists a deterministic $\mu' \in \mathbf{M}$ such that agent k's expected payoff from μ' is no less than his expected payoff from μ, both evaluated conditional on x, for all $k \in \mathbf{C} \cup \mathbf{S}$, with one strict inequality; μ is *ex-post efficient* if it cannot be ex-post dominated at any x.[13] Are modified serial dictatorship matching outcomes ex-post efficient? The following result gives a sufficient condition on preferences for a positive answer.

Proposition 1 *Let $w_{c,s}(x_s, q_s) = w_c(x_s, q_s)$ be decreasing in c, and $u_{1,s}(x) \neq 0$ for all $s \in \mathbf{S}$, all $x \in \mathbf{X}$. Then μ^{MSD} is ex-post efficient.*

The case in question represents natural environments where students rank unanimously as top colleges those that have a higher opportunity cost of admitting them, and therefore where it is more difficult to get in. Such a higher opportunity cost may be correlated to the faculty's research productivity, say, which would be distracted by the teaching and training of an additional student. A parametric example is provided by $w_{c,s}(x_s, q_s) = q_s - \gamma_c$, where the preference parameter $\gamma_c \in (0, 1)$ is the opportunity cost, and is increasing in c. In this case, even if simple serial dictatorship also delivers incentive posterior stable allocations, these allocations are not guaranteed to be ex-post efficient, as the following example shows.

Example 3 We take a symmetric affiliated environment with two colleges 1 and 2, and one student s. Each college receives three signals, i.e., $X_c = \{L, M, H\}$ with $L < M < H$, all c. The student prefers college 2 to college 1, and matching with any college over remaining unassigned, i.e., $v_2 > v_1 > v_s$. We suppose that: (i) college 1 prefers matching with the student unless one college has received the lowest signal and no college has received the highest signal; (ii) $u_{1,s}(x_1 = M, x_2 < H) < 0$; (iii) college 2 prefers matching with the student unless no college has received the highest signal; and (iv) $u_{2,s}(x_1 > L, x_2 = M) < 0$.

Let the deterministic allocations μ^1, μ^2 give rise to matchings[14]

$$m^1(s; x, \omega) = \begin{bmatrix} 2 & 2 & 2 \\ s & s & 1 \\ s & s & 1 \end{bmatrix}, \quad m^2(s; x, \omega) = \begin{bmatrix} 2 & 2 & 2 \\ s & 1 & 1 \\ s & s & 1 \end{bmatrix}$$

all ω. One can easily check that μ^1 is group stable and is obtained through simple serial dictatorship. However, it is not ex-post efficient, as it is dominated by $\mu^2 = \mu^{MSD}$. This shows that not only does modified serial dictatorship work where simple serial dictatorship may not, but when both work, the former has better efficiency properties.

[13] Note that for ex-post efficiency we allow improvements not to be strict. Also, because of the restrictions on alternative proposals (specifically, signal invariance), posterior efficiency does not automatically imply ex-post efficiency.

[14] Each cell identifies the student's match occurring when signals are (x_1, x_2); rows correspond to college 2's signals x_2, columns to college 1's signals x_1, going from L to H upward and rightward, respectively.

6 Conclusion

We consider two-sided many-to-one matching markets with interdependent valuations and imperfect information held by one side of the market. The other side has common and known preferences over potential mates. We propose a definition of group stable matching mechanisms for such environments and show that pairwise stability does not necessarily imply group stability. We constructively prove the existence of group stable matching mechanisms. We also identify precisely the restrictions on the ability of blocking coalitions to negotiate effectively that are necessary for obtaining general existence results.

In addition to traditional matching settings, our analysis also applies to situations where preference homogeneity for the uninformed side of the market arises because there is a single agent on that side. For instance, a principal (such as a department chair) may have to assign tasks to several agents (faculty members). The principal may have known preferences over the agents whom she ranks according to their ability or punctuality. But the agents may have interdependent values and private signals on the costs and benefits of completing the various tasks and they may update this information during the assignment process. Group stability in this context is the problem of designing a rule that aggregates information and assigns tasks to agents in a manner that is immune to misrepresentation of information as well as to objections raised by coalitions of agents (that may include possibly the principal herself) after the assignment is completed.

Appendix

In the proofs we use the following convenient notation. Given a mechanism μ, we denote by $m(k; x)$ the match of agent k realized at x for some $\omega \in [0, 1]$. We further omit the signal realization when convenient, and write $m(k)$ for such match realization. Also, with slight abuse of notation we let $\mu_{k,k'}(\hat{x})$ denote the probability that agent k is matched to agent k' under μ given reports $\hat{x}$.

Proof of Theorem 1 Given $x \in \mathbf{X}$, let $y_c = (x_{c+1}, \ldots, x_C)$ be the vector of signals $x_{c'}$ received by all $c' > c$, with $y_C = \varnothing$.

First, we construct modified serial dictatorship (MSD). Starting from college C and going down to college 1, for each college c let $\hat{k}_c$ be the number of students s such that $u_{c,s}(x_c, y_c) > 0$ and $v_c > v_{s,s}$, and that are not yet assigned to any college $c' > c$. Strictly order the students from best to worst, breaking ties among equal value students via uniform randomization. If $\hat{k}_c < k_c$, the mechanism assigns all the $\hat{k}_c$ students to college c leaving the remaining slots unassigned. Otherwise, the top k_c of the $\hat{k}_c$ students are assigned to college c. The procedure ends in finite time, as there are no cycles: the number of students and colleges is finite, and no student is evaluated for any college more than once. Let the constructed allocation rule be denoted μ^{MSD}. Observe that if $m(c; x)$ is a match realization for college c under μ^{MSD}, then $m(c; x) = m(c; x_c, y_c)$, all $c \in \mathbf{C}$. Let $u_{c,k_c}(\mathbf{F}_c)$ be the value of college c's k_c-th mate under μ^{MSD}, conditional on information $\mathbf{F}_c$.

We break the proof into three lemmas. Lemma 1 checks that μ^{MSD} cannot be blocked in principle under truthtelling, and Condition 2 in Definition 2 is satisfied. Lemma 2 shows that μ^{MSD} is interim incentive compatible, and Condition 1 is satisfied. Finally, Lemma 3 establishes Condition 3.

Lemma 1 *Mechanism μ^{MSD} satisfies Condition 2.*

Suppose that a coalition $\mathcal{C}$ blocks μ^{MSD} at $\hat{x}, x, m$, with $\hat{x} = x$, and $m = \mu^{MSD}(x, \omega)$, some ω. It must be that $\mathbf{C}' \neq \varnothing$: by assumption on preferences, a student would never want to drop a college and remain unmatched, which would be the only possible block in the absence of colleges in a coalition. Let $c^* = \min \mathbf{C}'$, the lowest-ranked college in the coalition (by A1.1, a well-defined concept). By condition (PS1) in Definition 1, and by A1.2 (college payoff separability across students), it must be that

$$u_{c^*,k}(\mathbf{F}_{c^*}) - u_{c^*,k'}(\mathbf{F}_{c^*}) > 0 \tag{1}$$

for some $k \in \mathbf{S}' \cup \{c^*\}$ and some $k' \in \mathbf{S} \cup \{c^*\}$, where $\mathbf{F}_{c^*}$ is the information private to college c^* after observing its own signal x_{c^*}, its own match $m(c^*)$, given acceptance rules $\alpha_{-c^*} = 1$ and given truthtelling $\widehat{x} = x$. There are three possible cases to be considered: (i) $k = c^*$, when college c^* just drops a student $k' \in \mathbf{S}$ without getting any student in exchange; in that case, the value of the last slot at college c^* must be $u_{c^*,k_{c^*}}(\mathbf{F}_{c^*}) = u_{c^*,c^*}(\mathbf{F}_{c^*}) = 0$; (ii) $k' = c^*$, when college c^* just gets a new student $k \in \mathbf{S}'$ without dropping any student in exchange –in that case, again the value of the last slot at college c^* is $u_{c^*,k_{c^*}}(\mathbf{F}_{c^*}) = u_{c^*,c^*}(\mathbf{F}_{c^*}) = 0$; (iii) $k \in \mathbf{S}'$ and $k' \in \mathbf{S}$, when college c^* is exchanging an assigned student k for a new student k'.

Consider first case (i). Since in this case $k' \in m(c^*)$, by construction of μ^{MSD} it must be that $u_{c^*,k'}(x_{c^*}, y_{c^*}) \geq 0$ for all (x_{c^*}, y_{c^*}) such that $k' \in m(c^*; x_{c^*}, y_{c^*}) = m(c^*)$. Conditional on truthtelling, the information $\alpha_{-c^*} = 1$ can only reveal $x_{c'}$ for some $c' \in \mathbf{C}'$, $c' > c^*$. Hence, as $u_{c^*,k'}(\mathbf{F}_{c^*})$ is an average across elements (x_{c^*}, y_{c^*}) as above, $u_{c^*,k'}(\mathbf{F}_{c^*}) \geq 0$, a contradiction.

Now consider case (ii). By A1.1 (homogeneous student preferences) and condition (PS2) in Definition 1, it must be that $k \in m(c)$ for some $c < c^*$, implying $c \notin \mathbf{C}'$, or $m(k) = k$. By construction of μ^{MSD} it is $u_{c^*,k}(x_{c^*}, y_{c^*}) \leq 0$ for all (x_{c^*}, y_{c^*}) such that $k \notin m(c^*; x_{c^*}, y_{c^*})$. The information $\mathbf{F}_{c^*}$ is a coarsening of the last two facts. Therefore, $u_{c^*,k'}(\mathbf{F}_{c^*})$ is an average across these states, implying $u_{c^*,k}(\mathbf{F}_{c^*}) \leq 0$, a contradiction.

Case (iii) is now a straightforward combination of the arguments in the previous two cases, and we then omit the details for brevity, ending the proof. □

Lemma 2 *Mechanism μ^{MSD} satisfies Condition 1.*

For any c, let x_c be the true signal for c, and $\widehat{x_c} \neq x_c$ be c's false report. Let

$$U_c(\mu; \hat{x}_c, x_c) = \sum_{y_c} U_c(\mu; \hat{x}_c, x_c, y_c) \Pr(y_c | x_c)$$

where $U_c(\mu; \hat{x}_c, x_c, y_c)$ is college c's expected payoff when this college reports $\hat{x}_c$, its signal is x_c, all other agents report truthfully, the match assigned by μ is not blocked in principle by any coalition, and we condition also on y_c. Note that by construction of μ^{MSD}, with its message college c cannot affect the set of its available mates, which is determined by y_c instead. Then, fix a student s and consider only x_c, y_c such that $u_{c',s}(x_{c',s}, y_{c',s}) \leq u_{c',k_{c'}}(x_{c'}, y_{c'})$ for all $c' > c$. It is easy to show that for every such x_c, y_c it is $u_{c,s}(\hat{x}_c, x_c, y_c) \leq u_{c,s}(x_c, x_c, y_c)$. Indeed, if the student is worth less than the cutoff k_c at x_c, y_c, lying is going to matter only if college c gets the student as a result; whereas if the student is worth at least as much as the cutoff k_c, lying matters only if college c drops the student as a result. In either case, lying clearly does not lead to a payoff improvement over μ^{MSD} conditional on x_c, y_c. Averaging across all y_c, we have $u_{c,s}(\hat{x}_c, x_c) \leq u_{c,s}(x_c, x_c)$ for all s. Since by A1.2 colleges' payoffs are separable in students, we conclude that $U_c(\mu; \hat{x}_c, x_c) \leq U_c(\mu; x_c, x_c)$ for all c, as wanted. □

Lemma 3 *Mechanism μ^{MSD} satisfies Condition 3.*

By contradiction, suppose not, and that μ is blocked in principle by some coalition $\mathcal{C}$ at $(\widehat{x}, x, m)$ satisfying $\widehat{x}_c \neq x_c$ for a unique $c \in \mathbf{C}'$, with $U_c(\mu'; \widehat{x}_c, x_c, \alpha_{\mathcal{C}}) > U_c(\mu; x_c, x_c)$. Since from Lemma 2 μ^{MSD} satisfies Condition 1, it follows that $\mu' \neq \mu$, i.e., the block is non-trivial.

Let $U_c(\mu'; \widehat{x}_c, x_c, \alpha_{\mathcal{C}}, y_c)$ be college c's expected payoff from blocking in principle when joining coalition $\mathcal{C}$ having reported message $\hat{x}_c$ when the signal is x_c, acceptance strategies in $\mathcal{C}$ are $\alpha_{\mathcal{C}}$ and the alternative is μ', and college c also observes y_c. Let $U_c(\mu; x_c, x_c, y_c)$ denote the similar payoff from truthtelling and accepting μ. Then, the assumed condition holds only if there exists y_c such that

$$U_c(\mu'; \widehat{x}_c, x_c, \alpha_{\mathcal{C}}, y_c) > U_c(\mu; x_c, x_c, y_c) \tag{2}$$

Notice first that for each c, given x_c, y_c and any message $\widehat{x}_c$, the observed outcome $m(c)$ contains no information for c that is not already contained in x_c, y_c. This follows from construction of μ^{MSD}.

Second, the blocking coalition $\mathcal{C} = (\mathbf{C}', \mathbf{S}')$—associated with college c having reported message $\hat{x}_c$ when the signal is x_c, acceptance strategies in $\mathcal{C}$ are $\alpha_{\mathcal{C}}$ and the alternative is μ', and college c observes y_c and the matching outcome $m(c)$ resulting from the report $\widehat{x}_c$—and everyone else reporting truthfully—must contain at least one other college c' with $c' < c$. This is because a blocking strategy where c is the singleton college in the coalition can only attract students lost to colleges $c' < c$ and drop those gained from $c' > c$ as a result of the message $\widehat{x}_c$. Further, it will not convey any information (via the acceptance strategies of students in the coalition) that c already does not know from x_c, y_c. Then, coalition $\mathcal{C}$ must involve a college $c' < c$ from whom c obtains information at the blocking stage via acceptance strategies.

Now consider $c^* = \min \mathbf{C}'$. Since c^* participates in the block under the presumption that c tells the truth, and because of the full support assumption, Lemma 1 concludes that c^* cannot be better off from joining this coalition, contradicting condition (PS1) in Definition 1. Thus, no such y_c can exist, (2) does not hold and Condition 3 must hold at all y_c, and therefore also on average, and this for all x_c and all c, concluding the proof. □

Proof of Theorem 2 (**R**1): We replace the strict inequalities in (PS1) and (PS2) in Definition 1 for all agents in $\mathcal{C}$ with weak inequalities for all but at least one agent, who should accept strictly. We call *group* * *stable* a mechanism that satisfies this modified notion. We consider the matching market given by Example 2. We start by excluding certain matching configurations. To begin, no group* stable mechanism μ^* can have $\mu^*_{s,s}(\hat{x}) > 0$ at any $\hat{x}$, or college 1 and the student will block at such $\hat{x}$ when $m(s) = s$, by (i). Next, it must be $\mu^*_{2,s}(L) = 0$, or college 2 will block alone if matched with the student at $\hat{x} = x = L$, by (ii). Similarly, $\mu^*_{1,s}(H) = 0$, or college 2 will object with the student upon observing $m(2) = 2$ at $\hat{x} = x = H$, again by (ii). Now we claim that $\mu^*_{2,s}(H) > 0$. For if not, the previous steps imply $\mu^*_{3,s}(H) = 1$. First suppose that $\mu^*_{1,s}(L) > 0$. Then, at $x = H$ college 2 can lie at the first stage and then block with the student when $m(s) = 1$. Under μ^* the student then gets $v_2 > v_1$, hence accepts if and only if $m(s) = 1$. College 2 gets $u_{2,s}(H) > 0$, by (ii); and in expectation before stage one, it gets $u_{2,s}(H)\mu^*_{1,s}(L) > 0$, zero being the expected payoff under μ^*. Hence, lying and blocking would be better than truthtelling and no rematching, a contradiction to Condition 3. If instead $\mu^*_{1,s}(L) = 0$, then by the previous steps $\mu^*_{3,s}(\hat{x}) = 1$ all $\hat{x}$, and college 3 would object alone, by (iii). Therefore, $\mu^*_{2,s}(H) > 0$.

Finally, we show that $\mu^*_{2,s}(H) > 0$ is impossible for a group* stable mechanism. Indeed, colleges 1 and 3 and the student would then block at $\hat{x} = x = H$ when $m(s) = 2$, with $m'(3) = s$ and the acceptance rules: $\alpha_1(m(1)) = 1$ if and only if $m(1) = 1$; $\alpha_3(m(3)) = 1$ always; $\alpha_s(m(s)) = 1$ if and only if $m(s) \neq 3$. The student's acceptance obviously satisfies (PS2) strictly. College 1 gets zero with μ^* or m' when $m(1) = 1$, and when $m(1) = s$ gets $0 < u_{1,s}(L)$, by (i), the latter being its payoff under μ^* , so α_1 satisfies (PS1) weakly. Using $\alpha_{-3} = 1$, college 3 knows that it is pivotal only when $x = H$. Indeed, while obviously college 3 is not worse off with m' if $m(3) = s$, when $m(3) = 3$ the college knows that it must be $m(1) = 1$ for them to be pivotal, but then $m(2) = s$, which can only occur when $x = H$. Its payoff from m' given α_1 is then $u_{3,s}(H) > 0$, by (iii), zero being its payoff under μ^*. Hence, college 3 is strictly better off with m', and μ^* can be blocked in principle. □

(**R**2): Alternatives μ' are allowed to be lotteries, and we call the resulting notion of stability *group*** *stability*. Again, we consider the matching market given by Example 2. Up to the last block considered in the proof under R1, restrictions on mechanism μ^* were obtained without recourse to the weak inequalities inherent to group* stability. Hence, allocation μ^{**} is group** stable only if it is subject to those same restrictions. To prove the claim it is then sufficient to show that the last block used in the proof under R1 can be modified and turned into a block in principle against $\mu^{**} = \mu^*$, the only candidate group** stable mechanism. To this aim, consider the coalition formed by colleges 1 and 3 , and the student, counterproposal μ' with $\mu'_{1,s} \in (0, 1)$, $\mu'_{1,s} + \mu'_{3,s} = 1$, and such that $\mu'_{1,s}v_1 + \mu'_{3,s}v_3 > v_2$, and acceptance rules $\alpha_c(m(c)) = 1$ if and only if $m(c) = c$, for $c = 1, 3$, and $\alpha_s(m(s)) = 1$ if and only if $m(s) \neq 3$. At $x = H$ and when $m(s) = 2$, this coalition will block μ^{**}. Indeed, when $m(1) = 1$ and college 1 is pivotal, under μ' college 1 gets the student with probability $\mu'_{1,s} > 0$, for a payoff of $[u_{1,s}(L)\Pr(L) + u_{1,s}(H)\Pr(H)]\mu'_{1,s} > 0$, zero being its payoff at $m(1) = 1$ under μ^{**}: college 1 is strictly better off with μ'. When $m(1) = s$, the college is strictly better off keeping the student for sure than only with probability $\mu'_{1,s}$. College 3 then is strictly better off with μ' only if $m(3) = 3$, as then it knows that it is pivotal only when $m(2) = s$ and its payoff is then $u_{3,s}(H)\mu'_{3,s} > 0$, by (iii), zero being its payoff under μ^{**}. □

(**R**3): We call the derived notion of stability *group**** *stability*. We consider the "replicated" Example 2. For each student $s_h = s, s'$, we can again apply the first steps in the proof under R1 to narrow down $\mu^{***}_{s_h}$ as a group*** stable allocation for student s_h , with the same restrictions. The remaining pairs $\left(\mu^{***}_s, \mu^{***}_{s'}\right)$ resulting from combinations of such assignments for students s, s' must then have in particular $\mu^{***}_{2,s}(H, x_{s'}) > 0$ and $\mu^{***}_{2,s}(x_s, H) > 0$, for any $x_s, x_{s'}$. We now show that at signals $x = (x_s, x_{s'}) = (H, H)$, the coalition of $\mathbf{C}' = \{1, 3\}$ and $\mathbf{S}' = \{s, s'\}$ can block in principle mechanisms μ^{***} by replicating the randomness of μ' under R2 via mixed strategy announcements as functions of payoff-irrelevant messages. In the interest of brevity, and to save on cumbersome yet obvious notation, we only sketch the main elements of the argument as follows, leaving the details to the reader.

Let the message space of mechanism μ' be the agent's type for colleges, and the student type augmented by a binary message $\{\xi^1_{s_h}, \xi^2_{s_h}\}$, for $s_h = s, s'$.

The map μ'_{s_h} is set as follows: when student s_h announces message $\xi^1_{s_h}$, the other student $s_{h'}$ is allocated to college 1; whereas when student s_h announces message $\xi^2_{s_h}$ the other student $s_{h'}$ is allocated to college 3, regardless of the message of other coalition members; for any other message sent by student s_h, student $s_{h'}$ is left unassigned. Notice that μ' does not depend on the messages sent by the colleges and that the allocation for any student does not depend on the messages sent by that student, but only on the messages sent by the other student.

For any student s_h, let $\mathbf{A}_{s_h}$ be the acceptance set for that student in the voting game, corresponding to types t_{s_h} where s_h has not been assigned to college 3 by μ^{***}. For $t_{s_h} \in \mathbf{A}_{s_h}$, let student s_h choose a mixed announcement strategy $\sigma_{s_h}(\xi^1_{s_h}|t_{s_h}) = p = 1 - \sigma_{s_h}(\xi^2_{s_h}|t_{s_h})$ where $pv_1 + (1-p)v_3 > v_2$, and otherwise let the announcement strategies for all parties be arbitrary. For colleges 1 and 3, let $\mathbf{A}_c$ correspond to the set of types for college c for which $\alpha_c = 1$ in each of the cases considered under R2 when such μ' was used. Observe that, with such a choice of μ', $\{\mathbf{A}_k\}_{k\in\mathcal{C}}$ and $\{\sigma_k\}_{k\in\mathcal{C}}$, from the perspective of each student s_h we have recreated the lottery μ' used under R2 by using the random messages sent by the other student. It then follows that self-selection and incentive compatibility conditions are satisfied, due to arguments identical to those used under R2. We conclude that no mechanism $\mu^{***} = \left(\mu^{***}_{s_1}, \mu^{***}_{s_2}\right)$ can have $\mu^{***}_{2,s_1}(H, H) > 0$ or $\mu^{***}_{2,s_1}(H, H) > 0$, a contradiction. Therefore, group*** stable mechanisms do not exist, ending the proof of the theorem. □

Proof of Proposition 1 Drop the superscript MSD from μ^{MSD}. Suppose μ is not ex-post efficient. Then at some $x \in \mathbf{X}$ there exists a deterministic $\mu' = m'$ that is feasible for $\mathcal{C} = (\mathbf{C}, \mathbf{S})$ and such that $U_k(m'(k) - m(k)|x) \geq 0$ for all $k \in \mathbf{C} \cup \mathbf{S}$, with one strict inequality.

Because of dominance for students, no college can just drop any student. Because some agent must be strictly better off, the assignment must change at least for a student, hence for a college. Let $\bar{c} \geq 1$ be the highest such college. College $\bar{c}$ must receive a student s either assigned to college $c' < \bar{c}$ under μ, or unassigned by μ.

If the second case, $u_{1,s}(x) \leq 0$ by construction of μ –and $u_{1,s}(x) < 0$ by assumption.

Since $w_{\bar{c}} \leq w_1$, it must be that $u_{\bar{c},s}(x) < 0$. If college $\bar{c}$ then gives up a student s'' with $u_{\bar{c},s''}(x) < 0$, it must be to a college $c'' > \bar{c}$, by A1.1, contradicting the fact that $\bar{c}$ is the highest college for which the assignment changes.

Hence, college $\bar{c}$ must receive a student s' assigned to college $c' < \bar{c}$ under μ, also the first case (in which case, without loss of generality we can assume $u_{\bar{c},s}(x) \geq 0$, and hereafter let $s' = s$). Since $w_{\bar{c}} < w_{c'}$, it is $u_{c',s'}(x) > 0$ and college c' must receive student s'' such that $u_{c',s''}(x) > 0$ from $c'' < c'$. Indeed, if student s'' was unassigned, then $u_{1,s''}(x) < 0$ by construction of μ and by assumption, and by assumption on w it would be $u_{c',s''}(x) < 0$, a contradiction. Continuing this logic, we arrive at college $c^n = 1$ and at a student s^n with $u_{1,s^n}(x) > 0$ where s^n was unassigned. Since for all s unassigned under μ we have $u_{1,s}(x) < 0$, we have a contradiction. □

References

Afacan M (2012) Group robust stability in matching markets. Games Econ Behav 74:394–398

Chakraborty A, Citanna A, Ostrovsky M (2010) Two-sided matching with interdependent values. J Econ Theory 145:85–105

Chade H (2006) Matching with noise and the acceptance curse. J Econ Theory 129:81–113

Dutta B, Vohra R (2005) Incomplete information, credibility and the core. Math Soc Sci 50:148–165

Ely JC, Siegel R (2013) Adverse selection and unraveling in common-value labor markets. Theor Econ 8:801–827

Forges F (1994) Posterior efficiency. Games Econ Behav 6:238–261

Forges F, Mertens JF, Vohra R (2001) The ex ante incentive compatible core in the absence of wealth effects. Econometrica 70:1865–92

Forges F, Minelli E, Vohra R (2001) Incentives and the core of an exchange economy: a survey. J Math Econ 38:1–41

Hoppe H, Moldovanu B, Sela A (2009) The theory of assortative matching based on costly signals. Rev Econ Stud 76:253–281

Kojima F (2011) Robust stability in matching markets. Theor Econ 6:257–267

Myerson R (1991) Game theory. Harvard University Press, Cambridge, MA

Myerson R (1982) Optimal coordination mechanisms in generalized principal-agent problems. J Math Econ 10:67–81

Neeman Z, Pavlov G (2013) Ex post renegotiation proof mechanism design. J Econ Theory 148:473–501

Roth A (1989) Two-sided matching with incomplete information about others' preferences. Games Econ Behav 1:191–209

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

Shi S (2002) A directed search model of inequality with heterogeneous skills and skill-biased technology. Rev Econ Stud 69:467–491

Vohra R (1999) Incomplete information, incentive compatibility, and the core. J Econ Theory 86:123–147

Wilson R (1978) Information, efficiency, and the core of an economy. Econometrica 46:807–16