Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 157 (2015) 1–27
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

A one-shot deviation principle for stability in matching problems

Jonathan Newton [a,1], Ryoji Sawa [b,*]

[a] School of Economics, University of Sydney, Australia
[b] Center for Cultural Research and Studies, University of Aizu, Japan

Received 11 June 2014; final version received 13 November 2014; accepted 26 November 2014

Available online 3 December 2014

Abstract

This paper considers marriage problems, roommate problems with nonempty core, and college admissions problems with responsive preferences. All stochastically stable matchings are shown to be contained in the set of matchings which are most robust to one-shot deviation.

JEL classification: C71; C72; C73; C78; D71

Keywords: Learning; Stochastic stability; Matching; Marriage; College admissions

1. Introduction

Partnerships fail. Marriages break down, friendships rupture, your gym buddy stops training. When partnerships break down, new partnerships are forged in the aftermath, until an equilibrium, or something close to an equilibrium, is again reached. The reasons that partnerships can break down are many: often human imperfection and the vicissitudes of fate play a role. Errors, mishaps or misbehavior on the part of one of the partners can contribute to the decline of a part-

* Corresponding author at: Tsuruga, Ikki-machi, Aizu-Wakamatsu City, Fukushima, 965-8580 Japan.
E-mail addresses: jonathan.newton@sydney.edu.au (J. Newton), rsawa@u-aizu.ac.jp (R. Sawa).
1 This work was completed while the author was supported by a Discovery Early Career Researcher Award funded by the Australian Research Council.

http://dx.doi.org/10.1016/j.jet.2014.11.015

nership. However, partnerships are not all alike. Some partnerships are strong, some are weak. Some partnerships are easily substitutable, others less so. It is not just the partnerships themselves that can be more or less robust. Due to the interrelationship of different partnerships, networks of partnerships also display robustness characteristics which depend on the robustness of their constituent pairings. This paper analyses such settings in the context of the well known marriage problem of Gale and Shapley [16] as described in Jackson and Watts [19]. We show that for standard matching dynamics, perturbed by an error process, any stochastically stable matching is contained in the class of matchings which are most robust to one-shot deviation. For the logit choice rule, this class corresponds to a non-transferable utility version of the least-core as described in Maschler, Peleg and Shapley [25]. The results extend to one-sided matching markets (roommate problems) and to many-to-one matchings (college admissions problems).

Similarly to the related papers of Jackson and Watts [19] and Klaus, Klijn and Walzl [21], players occasionally make mistakes in a dynamic model of partnership formation. Mistakes involve a player leaving an existing partner or matching with a new partner in such a way that his payoff is reduced. Mistakes can be fatal to a partnership and can drive the dynamic process of partnership formation to a new equilibrium. Jackson and Watts [19] and Klaus et al. [21] derive results for stochastic stability of marriage and roommate problems under *uniform* mistake probabilities: every possible mistake has the same order of probability of occurring. They find that all stable matchings are *stochastically stable* in the sense of Kandori, Mailath and Rob [20]; Young [40].[2] When mistakes are rare, in the long run the process will spend almost all of its time at stochastically stable matchings.

The current paper addresses a large class of alternative mistake models, including payoff dependent models such as logit choice [6] and probit choice [11,26], for which the decision rule depends explicitly on cardinal preferences.[3] Due to differing strengths of partnerships, the authors of the current paper believe cardinal preferences to be a natural assumption in matching models. Moreover, abstraction away from cardinal preferences, or the choice of a dynamic which is insensitive to such preferences, is not without loss when it comes to applying a concept such as stochastic stability.[4] As pointed out by Bergin and Lipman [3], the identity of stochastically stable states depends on the mistake model. Therefore, results which are applicable across a broad range of mistake models are of particular interest.

In this paper, it is no longer the case that all stable matchings are stochastically stable. Given the large class of mistake models considered, this is unremarkable [3]. What is remarkable is that for all of these models there exists a simple local property that must be satisfied by any stochastically stable matching. Specifically, this paper shows that stochastically stable matchings must be contained in the set of stable matchings which are most robust to one-shot deviation, that is the set of stable matchings at which the most probable mistake is not more probable than the most probable mistake at any other stable matching. When this set is a singleton, as at least

[2] In the language of Noldeke and Samuelson [30], they find that all the equilibria are part of a single *mutation connected component*. This fact means that all of the stable partnership networks in their setting are stochastically stable under uniform mistakes.

[3] The relation of such rules to uniform mistake models can be thought of as similar to the relation between the static concepts of Proper Equilibrium [27] and Trembling Hand Perfect Equilibrium [38]. In the former, mistakes associated with larger payoff losses are less likely, whereas in the latter there is no difference.

[4] It is not uncommon to assume cardinal preferences in the literature on matching problems. Abdulkadiroğlu, Che and Yasuda [1] discuss that a mechanism sensitive to cardinal preferences may achieve a Pareto-superior matching to one obtained by the deferred acceptance mechanism of Gale and Shapley [16]. In an experimental study of decentralized matching, Echenique and Yariv [12] find that cardinal preferences have a clear effect on which stable matching is selected.

one stochastically stable state always exists, its unique member must be the unique stochastically stable state. These results hold for marriage problems, roommate problems with nonempty core, and college admissions problems with responsive preferences.

Our result is surprising because stochastic stability is a globally determined property: existing characterizations [20,40] and partial characterizations [13] depend on transition paths between *all* of the stable states. Computing probabilities for all such transition paths can be cumbersome. In contrast, the set of stable matchings which are most robust to one-shot deviation is defined solely by reference to local properties of the stable matchings. To compare with another contribution to the partial characterization literature, Ellison [13] provided a globally determined sufficient condition for stochastic stability in any (finite) problem, whereas we provide a locally determined necessary condition for stochastic stability in matching problems.

For the logit choice rule, the most probable mistake at a stable matching is the deviation which causes the lowest payoff loss to the deviating players. The set of stable matchings which are most robust to one-shot deviation maximize this lowest possible payoff loss. This set is equal to a non-transferable utility version of the least-core as described in Maschler et al. [25]. That is, there is a connection between perturbed adaptive dynamics, matching problems, robustness to one-shot deviations and a well known concept in cooperative game theory.[5]

There is a growing literature which looks at equilibrium selection in matching problems [5, 7,12,31]. Typically, these papers use simulation[6] or experimental evidence to generate a distribution over absorbing states reached by a dynamic process *without* mistakes, conditional on the process being started at some initial matching. In contrast to these papers, our results are independent of the initial matching and the probabilities with which any players are chosen to better respond. Moreover, the results in the current paper are analytical.[7] The papers cited above consider short run behavior given some initial condition. In contrast, the current paper models the long run.

A useful literature from the perspective of the current paper is the *paths to stability* literature in matching problems with non-transferable utility. This focuses on convergence to core allocations in situations where the payoff for an individual depends only on his partner [10,35]. Another related literature is the literature on *convergence to the core* in cooperative games [2,14,17,29]. A branch of this literature has recently explicitly focused on the case in which all relevant coalitions are pairs – the transferable utility equivalent of the marriage problem, otherwise known as the assignment problem [4,9,22,28,39]. Of particular note is the work of Nax and Pradelski [28], who adapt the results of the current paper to obtain selection within the core of the assignment game.

In contrast to the literature on paths to stability, the processes in the current paper go on forever. Much previous literature on matching in economics considers algorithms that reach a final matching (school choice, hospital-intern, kidney donation). However, in some situations the possibility of extensive rematching can persist indefinitely. An example is within-firm collaborative working arrangements such as "pair programming". Another example is bilateral business re-

---

[5] We thank Bary Pradelski and Heinrich Nax for bringing this connection to our attention. Nax and Pradelski [28] adapt the methods of the current paper to give least-core selection in assignment problems.

[6] There exist software tools that support such computations. Biró and Norman [5] use the PRISM model checker [24] to obtain results.

[7] Boudreau [8] writes of the prior approach: *"Calculating the probability of each stable outcome for a given market under the randomized tâtonnement process is extremely difficult due to the tremendous number of paths that can be involved... Loops in the process mean that a closed form solution is virtually impossible to obtain."*

lationships where preferences depend on things other than price. For example, a manufacturer may choose suppliers based on locational factors such as tariffs, political risk or quality of available human capital. These things are hard to change and thus make such matching problems more like NTU problems than TU problems. Even in centralized settings, where some central authority determines a matching, the central authority may not have the power to prevent further rematching after the determined matching have been implemented. Consequently, they may prefer to implement matchings that are relatively robust to stochastic choice behavior.

The paper is organized as follows. Section 2 gives the model and some relevant concepts from the literature. Section 3 gives the main results for marriage problems. Section 4 applies our results to marriage problems under differing choice rules. Sections 5 and 6 extend our main result to many-to-one matching problems and to roommate problems respectively.

2. Model

2.1. The marriage problem

We follow the description of the marriage problem in Jackson and Watts [19]. There is a set of players, N, which is divided into a set of men, $M = \{m_1, \dots, m_k\}$, and a set of women, $W = \{w_1, \dots, w_l\}$. An undirected network g is a set of edges $ij \in g$, each comprising a pair of players $i, j \in N$, $i \neq j$, such that $ij \in g \Leftrightarrow ji \in g$. Let $\mathcal{G}$ denote the set of all undirected networks on N. Let $g(i) = \{j : ij \in g\}$ denote the set of players linked to player i in network g. $g(i) = \varnothing$ means that i is single in g. The set of matchings in the marriage problem, G, is the set of undirected networks in which each woman is linked to at most one man, and each man is linked to at most one woman:

$$G = \big\{g \in \mathcal{G} : (\forall ij \in g, i \in M \Leftrightarrow j \in W),\ \big(\forall i \in N, \big|g(i)\big| \leq 1\big)\big\}.$$

In a slight abuse of notation, we sometimes write $g(i) = j$ for $g(i) = \{j\}$. Let $\mu = \{(i, j) : (\exists g \in G : ij \in g)\}$ be the set of pairs of players between whom a link can potentially exist.

The vector of utilities obtained from network g by the players is given by $u : G \to \mathbb{R}^{|N|}$. Player i obtains utility $u_i(g)$ from network g, and this utility depends only on the match of i. That is, for each i, $u_i(g) = u_i(g')$ if $g(i) = g'(i)$. We assume that players are never indifferent between two potential matches: $g(i) \neq g'(i)$ implies that $u_i(g) \neq u_i(g')$. Therefore, if $g(i) \neq \varnothing$, then $u_i(g) = u_i(\{ig(i)\})$, and if $g(i) = \varnothing$, then $u_i(g) = u_i(\varnothing)$. Define $g - ij := g \setminus \{ij\}$ as the network g with the edge ij removed if it exists in g. Similarly, define $g + ij := (g \setminus \{kl : k = i, l \in g(i) \text{ or } k = j, l \in g(j)\}) \cup \{ij\}$ as the network g with the edge ij added and any existing edges exiting i and j removed.

Definition 2.1. A matching $g \in G$ is stable if:

(i) $\forall ij \in g, u_i(g) > u_i(g - ij)$.
(ii) $\nexists i \in M, j \in W : u_i(g + ij) > u_i(g)$ and $u_j(g + ij) > u_j(g)$.

We denote the set of stable matchings by $\mathfrak{C}$. The set of stable matchings corresponds to the core of the problem: the set of matchings from which no subset of players can improve their payoffs by removing and adding edges in a coordinated manner.

2.2. Unperturbed blocking dynamic

We describe a class of unperturbed blocking dynamics.[8] Let g^t be the network in period t. At the beginning of period $t+1$, a pair of players (i, j) is selected at random according to a distribution $F_{g^t}(.)$ with full support on μ. Let g^{t+1} be determined as follows:

(i) If $g^t(i) = j$ and either $u_i(g^t - ij) > u_i(g^t)$ or $u_j(g^t - ij) > u_j(g^t)$, then, with some probability greater than zero, set $g^{t+1} = g^t - ij$.
(ii) If $g^t(i) \neq j$, $u_i(g^t + ij) > u_i(g^t)$ and $u_j(g^t + ij) > u_j(g^t)$, then, with some probability greater than zero, set $g^{t+1} = g^t + ij$.
(iii) $g^{t+1} = g^t$ otherwise.

In the terminology of matching problems, a pair $(i, j) \in \mu$ *blocks* a matching g if they prefer one another to their partners in g. Denote the transition probabilities of a given unperturbed blocking dynamic by $P_0(., .)$. That is, $P_0(g, g')$ is the probability that $g^{t+1} = g'$, given that $g^t = g$.

2.3. Perturbed blocking dynamic

Players meet and will usually take the myopically optimal action, whether that is to stay with their current partner, dissolve an existing partnership, or create a new partnership. However, from time to time, players make mistakes and take actions which reduce their payoffs, whether it be leaving or creating a partnership. That is, a pair selected by the dynamic will sever an existing beneficial link, or create a link which is worse than the status quo for at least one of the players involved. We consider families of perturbed blocking dynamics, with transition probabilities $P_\eta(., .)$, indexed by a parameter $\eta \in (0, \bar{\eta})$. The family $\{P_\eta\}_{\eta \in (0,\bar{\eta})}$ is assumed to satisfy the following conditions.

Assumption 1 *(Conditions on the perturbed dynamic).*

(i) $P_\eta \xrightarrow{\eta \to 0} P_0$, where P_0 are the transition probabilities for some unperturbed blocking dynamic as described in Section 2.2.
(ii) For $\eta > 0$, the chain induced by P_η is irreducible.
(iii) P_η vary continuously in η.
(iv) If, for $g \neq g'$, $P_0(g, g') = 0$, $P_{\hat{\eta}}(g, g') > 0$ for some $\hat{\eta} > 0$, then $\lim_{\eta \to 0} -\eta \log P_\eta(g, g') = c$ for some $c > 0$.
(v) For any $\eta \geq 0$, $P_\eta(g, g') > 0$ implies $g' = g + ij$ or $g' = g - ij$ for some $(i, j) \in \mu$.

Condition (i) merely states that the family of perturbed dynamics corresponds to an unperturbed dynamic. Conditions (ii), (iii), (iv) restrict the process to *weakly regular* Markov chains. A broad class of strategy revision rules falls into this category. Examples include best response with mutations, the logit choice rule, pairwise comparison rules, and the probit choice rule (see Sandholm [36]). Condition (v) means that transitions always involve a single pair of players getting together or splitting up. This restriction is needed to eliminate the possibility of two couples

[8] Our unperturbed dynamic is essentially the same as those of Roth and Vande Vate [35], Jackson and Watts [19] and Klaus et al. [21].

separating (or getting together) at the same point in time with a higher (order of) probability than either one of the couples acting alone.

As a chain with $\eta > 0$ is irreducible, there exists a unique stationary distribution π_η. For convenience, we assume the following.[9]

Assumption 2 *(Existence of limit).*

$$\pi_0 := \lim_{\eta \to 0} \pi_\eta \quad \text{exists.}$$

A matching g is *stochastically stable* if $\pi_0(g) > 0$. We denote the set of stochastically stable states by SS.

Definition 2.2.

$$SS := \{g \in G : \pi_0(g) > 0\}.$$

All stochastically stable matchings belong to recurrent classes of the unperturbed process [40] and from any matching there exists a finite sequence of transitions under the unperturbed process that culminates in a stable matching being reached [19,35]. Therefore, the only recurrent classes of the unperturbed process are the individual stable states, i.e. $SS \subseteq \mathfrak{C}$. The identity of the stochastically stable matchings is important, as for small error probabilities the process will spend almost all of the time at these matchings.

2.4. *Costs of transitions*

The identity of stochastically stable states depends on the transition probabilities of the process. To measure the limiting relative magnitude of these probabilities, a cost function is defined as follows.

Definition 2.3. The 1-step cost of the process moving from g to g' is defined as:

$$c(g, g') := \lim_{\eta \to 0} -\eta \log P_\eta(g, g'), \tag{1}$$

adopting the convention that $-\log 0 = \infty$.

$c(g, g')$ is the exponential decay rate of the transition probability from g to g'. The rarer a transition, the higher its cost. Impossible transitions have infinite cost. Note that for $g \notin \mathfrak{C}$, there is a zero cost transition from g. This is because there is some $g' \neq g$, such that $P_\eta(g, g')$ does not approach zero as $\eta \to 0$. We are also interested in the overall cost of moving between g and g', even if many steps are required. Let the t-step transition probabilities be given by $P^t_\eta(g, g') \equiv P(g^t = g' | g^0 = g, P_\eta(.,.))$.

Definition 2.4. The overall cost of the process moving from g to g' is defined as:

$$C(g, g') := \min_{t \in \mathbb{N}} \lim_{\eta \to 0} -\eta \log P^t_\eta(g, g'). \tag{2}$$

[9] This condition could be avoided if stochastically stable states were defined as states for which $\pi_\eta(.) \nrightarrow 0$ as $\eta \to 0$.

We make one further assumption: we rule out other-regarding mistake probabilities. That is, the cost of a mistake by a pair (i, j) is independent of the current matching of every player other than i and j. Given that the unperturbed dynamic is self-regarding, this seems a reasonable restriction.

Assumption 3 *(Self-regarding mistake probabilities).* If $g(i) = g'(i)$ and $g(j) = g'(j)$, then $c(g, g - ij) = c(g', g' - ij)$ and $c(g, g + ij) = c(g', g' + ij)$.

A *spanning tree* rooted at $g^* \in \mathfrak{C}$ is a directed graph over the set $\mathfrak{C}$ such that every $g \in \mathfrak{C}$ other than g^* has exactly one exiting edge, and the graph has no cycles (implying that g^* has no exiting edges). The *cost* of a spanning tree is the sum of the costs of its edges given by $C(.,.)$. A *minimum cost spanning tree* is a spanning tree whose cost is lower than or equal to the cost of any other spanning tree. A state $g^* \in \mathfrak{C}$ is stochastically stable only if there exists a minimum cost spanning tree rooted at g^* [40].[10] Finding minimum cost spanning trees can be difficult.[11] The principal contribution of the current paper is to show that the root of any minimum cost spanning tree, and hence any stochastically stable matching, must be in the set of matchings which are most robust to one-shot deviation. We call a transition $g \to g'$ from a matching $g \in G$ the *least cost deviation* from g if it has the lowest cost of all possible 1-step transitions from g.

Definition 2.5. Denote the set of possible least cost deviations from $g \in G$ by:

$$\mathrm{L}(g) := \arg\min_{g' \neq g} c\big(g, g'\big)$$

and the set of pairs of players involved in least cost deviations from $g \in G$ by:

$$N_{\mathrm{L}}(g) := \big\{(i, j) \in M \times W : \exists g' \in \mathrm{L}(g) : g' = g - ij \text{ or } g' = g + ij\big\}$$

$c_{\mathrm{L}}(g)$ will be used to denote the cost of the least cost deviation from g.[12]

$$c_{\mathrm{L}}(g) := \min_{g' \neq g} c\big(g, g'\big).$$

We use the word *deviation* as we shall be interested in the application of these concepts to $g \in \mathfrak{C}$.

3. Stochastically stable matchings

Define OS, the set of matchings which are most robust to one-shot deviation:

---

10 For many dynamics, ‘only if’ can be replaced by ‘if and only if’. See Sandholm [36] for details.

11 The same applies to radius-(modified)coradius methods [13]. The major difficulty lies in calculating $C(g, g')$, $g, g' \in \mathfrak{C}$. This is equivalent to solving a shortest path problem on a directed graph on the state space, which for the current problem is G. Although this can be solved in polynomial (in $|G|$) time (e.g. Fredman and Tarjan [15]), $|G|$ itself increases faster than the factorial of $\min\{|M|, |W|\}$. In contrast, the number of possible deviations from any given stable state only increases in $|M| \cdot |W|$.

12 This differs from the concept of the radius of a stable state $g \in \mathfrak{C}$ (Ellison [13], citing a no longer extant working paper of Evans, 1993). The radius is defined as $R(g) = \min_{g' \in \mathfrak{C} \setminus \{g\}} C(g, g')$ and requires a different stable state to be reached by the process. It turns out that in the problems considered in the current paper $c_{\mathrm{L}}(g) = R(g)$ for all stable matchings outside of a specific set, but this does not follow from the definitions.

$$OS = \left\{ g \in G : c_{\mathrm{L}}(g) = \max_{g' \in G} c_{\mathrm{L}}\left(g'\right) \right\}.$$

As $c_{\mathrm{L}}(g)$ is strictly positive only for $g \in \mathfrak{C}$, it must be that $OS \subseteq \mathfrak{C}$. We will show that OS contains SS: a stochastically stable matching must be comparatively robust against one-shot deviation. If OS contains only one matching, then that matching must be uniquely stochastically stable.

Klaus et al. [21] show that a single mistake suffices to move from any $g \in \mathfrak{C}$ to some other $g' \in \mathfrak{C}$. We show that the least cost deviation from a stable matching $g \notin OS$ is enough to escape from its basin of attraction, and that the unperturbed dynamic can subsequently lead the process closer to $OS \subseteq \mathfrak{C}$. This result is proved in Lemma 3.3, from which the main theorem is proven using a minimal cost spanning tree argument.

The following lemma, which assists in the proof of Lemma 3.3, shows that if a pair is involved in a least cost deviation from a stable matching $g \notin OS$, then the players forming the pair do not both have the same current partner as in some matching within OS. As any player who is single at some stable matching is single at every stable matching (Theorem 2.22 of Roth and Sotomayor [34]), this further implies that the least cost deviation from g cannot involve two single players forming a partnership.

Lemma 3.1. *Suppose that $g \in \mathfrak{C}$ and $g \notin OS$. If $(i, j) \in N_{\mathrm{L}}(g)$, then for all $g^* \in OS$, $g(i) \neq g^*(i)$ and/or $g(j) \neq g^*(j)$.*[13]

Proof. Let $g^* \in OS$. Suppose $g(i) = g^*(i)$ and $g(j) = g^*(j)$. If $g(i) = j$, then $c_{\mathrm{L}}(g^*) \leq c(g^*, g^* - ij) = c(g, g - ij) = c_{\mathrm{L}}(g)$. If $g(i) \neq j$, then $c_{\mathrm{L}}(g^*) \leq c(g^*, g^* + ij) = c(g, g + ij) = c_{\mathrm{L}}(g)$. Therefore $g \in OS$, which contradicts our premise. □

We now present the key lemma, which asserts that following the least cost deviation from any stable matching $g \notin OS$, the unperturbed dynamic can move to another stable matching which is strictly closer to OS than the initial matching. First, we define an index m which measures the similarity between matchings.

Definition 3.2. $m(g, g')$ is the number of players who have the same partner in g and g'.

$$m\left(g, g'\right) := \left|\left\{ i \in N : g(i) = g'(i) \right\}\right|.$$

Lemma 3.3 *(Getting Closer Lemma). Let $g^* \in OS$. Suppose that $g \in \mathfrak{C}$ and $g \notin OS$. Let $g_1 \in \mathrm{L}(g)$. Then, $\exists g' \in \mathfrak{C}$, $t \in \mathbb{N}_+$, such that $m(g^*, g') > m(g^*, g)$ and $P_0^t(g_1, g') > 0$.*

The proof of Lemma 3.3 is given in Appendix A. Using Fig. 1 to illustrate our argument, we here emphasize how our result is stronger than existing results in the literature, such as Lemma 5 of Klaus et al. [21] and a similar claim in Diamantoudi et al. [10]. These results show that, starting from any given unstable matching, it is possible, under the unperturbed dynamic, to reach a stable matching $g' \in \mathfrak{C}$ which is strictly closer to a target stable matching g^* than the initial unstable matching is to g^*. In Fig. 1, this result corresponds to the existence of a zero cost path from g_1 and g_2 to some stable matching which is closer to g^*. From g_1, it may be the case that any such path reaches g. The deviation from g to g_1 may not lead to a stable matching closer to g^*. Such a result suffices for the subsequent stochastic stability arguments of Klaus et al. [21], as

[13] We use the phrase ‘case A and/or case B’ to indicate that one or more of the cases may occur.

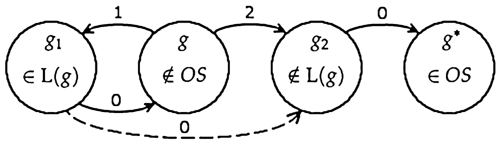

Fig. 1. g, g^* are stable matchings. g_1, g_2 are unstable matchings. Distance on the page represents distance under $m(.,.)$. Directed edges represent transitions and the numbers above the edges their respective transition costs.

all mistakes have the same cost in their model, so from g it is possible to 'choose' a desirable deviation such as the one to g_2 in Fig. 1. From g_2, g^* can be reached at no further cost. This option is not open to us. It may be that the least cost deviation from g moves *away* from g^*, such as is the case in the deviation from g to g_1 in the figure. We prove, using the structure of stable matchings and Lemma 3.1, that if $g \notin OS$ and $g^* \in OS$, then there exists a path (the dashed line in the figure) from g_1 that circumvents g, reaching an unstable state such as g_2, which is at least as close to g^* as g is to g^*. The application of previous results (generalized to the many to one case by Lemma 5.5 of the current paper) completes the argument. The least cost deviation from g suffices to move the process to a stable matching which is closer to $g^* \in OS$. The example in Section 4.2 will further illustrate these arguments.

Note that Lemma 3.3 and its many-to-one equivalent later in the paper can be understood as 'paths to stability' results which are stronger than the existing results in the literature. They allow us to say more than could previously be said about which stable states can be reached from different starting points. The knowledge of these paths gained from the lemma is exactly what is required to prove the main theorem.

Theorem 3.4. *SS* $\subseteq$ *OS*.

The formal proof is in Appendix A. In brief, any stochastically stable matching must be the root of a minimum cost spanning tree. If a tree is rooted at some $g \in \mathfrak{C}$, $g \notin OS$, then Lemma 3.3 can be used to build another tree rooted at some matching in *OS*. Take any $g^* \in OS$. Starting at g, use Lemma 3.3 to add edges between stable matchings which get progressively closer to g^*, stopping when some matching in *OS* is reached. We obtain a sequence $(g = g_1, \dots, g_L \in OS)$ with edges between g_i and g_{i+1} for $i = 1, \dots, L-1$. Each of these new edges has the cost of a lowest cost deviation, $C(g_i, g_{i+1}) = c_{\mathrm{L}}(g_i)$. Deleting the edge exiting g_L, we are left with a tree rooted at g_L. As $g \notin OS$ and $g_L \in OS$, the cost of the new edge exiting g must be lower than the cost of the deleted edge which exited g_L. So the tree rooted at g_L has a lower total cost than the total cost of the tree rooted at g. Therefore no tree rooted at g can be a minimum cost spanning tree. That is, $g \notin SS$.

Remark 3.5. Consider the special case of uniform mistake probabilities, that is when there exists $a \in \mathbb{R}$ such that for all g, $g' = g - ij$ or $g' = g + ij$ for some $i \in M$, $j \in W$, $c(g, g') > 0$ implies $c(g, g') = a$. It follows immediately from the proof of Theorem 3.4 that $SS = OS$. The result of Jackson and Watts [19] and Klaus et al. [21] is recovered.

Finally, we note that the proof of Theorem 3.4 extends to give a bound on convergence times.

Remark 3.6. It follows immediately from the proof of Theorem 3.4 that the modified-coradius (see Ellison [13]) of OS equals $\max_{g \notin OS} c_{\mathrm{L}}(g)$ and that therefore, starting from any matching, the expected hitting time of OS is $O(e^{\frac{1}{\eta}\max_{g \notin OS} c_{\mathrm{L}}(g)})$.

So, $SS \subseteq OS$. This is important, as the set OS is defined solely by reference to local properties of the stable matchings. Stochastically stable matchings must be matchings which are most robust to one-shot deviation. If OS is a singleton, then the unique stochastically stable state can be determined solely by looking at the lowest cost one-shot deviation from stable states: there is no need to resort to minimal cost spanning trees or to radius-coradius methods. If OS is not a singleton, then Theorem 3.4 assists in the use of such techniques by eliminating all states in $\mathfrak{C} \setminus OS$ as candidates for stochastic stability.[14]

4. Examples

In this section we apply the one-shot deviation principle to study stochastic stability under commonly used choice rules. In Section 4.1, we consider a dynamic induced by the logit choice rule and link OS to the notion of the *least-core* proposed by Maschler et al. [25]. Thus, our one-shot deviation principle combined with logit choice provides an evolutionary foundation for the least-core. In Section 4.2, we provide a comparison of dynamics induced by three leading choice rules, the uniform mistake, the logit choice, and the probit choice.

4.1. The logit choice rule

At the beginning of period $t+1$, a pair of players (i, j) is selected at random according to a distribution $F_{g^t}(.)$ with full support on μ. g^{t+1} is determined as follows:

(i) If $g^t(i) = j$, then $g^{t+1} = g^t - ij$ with probability

$$1 - \prod_{k \in \{i,j\}} \frac{e^{\frac{1}{\eta}u_k(g^t)}}{e^{\frac{1}{\eta}u_k(g^t)} + e^{\frac{1}{\eta}u_k(g^t - ij)}}.$$

That is, each of i and j chooses to cut or retain the link ij with probabilities given by the logit choice rule, and unless both players choose to retain the link, it will be cut.

(ii) If $g^t(i) \neq j$, then $g^{t+1} = g^t + ij$ with probability

$$\prod_{k \in \{i,j\}} \frac{e^{\frac{1}{\eta}u_k(g^t + ij)}}{e^{\frac{1}{\eta}u_k(g^t)} + e^{\frac{1}{\eta}u_k(g^t + ij)}}.$$

That is, i and j each agree to leave their existing partner and form a new link ij with probability given by the logit choice rule. Both i and j must agree for a new partnership to be formed.

(iii) $g^{t+1} = g^t$ otherwise.

---

[14] When using spanning tree methods, the number of relevant trees will decrease by a factor of $|OS|/|\mathfrak{C}|$. Furthermore, Lemma 3.3 and the results used in its proof will considerably facilitate calculation of $C(g, g')$, $g, g' \in \mathfrak{C}$ (see Footnote 11).

Under the logit choice rule, transition probabilities are sensitive to the amount by which cardinal utility is reduced. The sum of negative changes in revising players' payoffs for transition $g \to g'$ is the cost of $g \to g'$ [37]. If the easiest transition at matching g is for two players to form a partnership, then:

$$c_L(g) = \min_{ij \notin g}\left[\max\left\{u_i(g) - u_i(g+ij), 0\right\} + \max\left\{u_j(g) - u_j(g+ij), 0\right\}\right], \tag{3}$$

whereas if the easiest transition at matching g is for a player to dissolve an existing partnership, then:

$$c_L(g) = \min_{i:g(i)\neq\varnothing}\left[\max\left\{u_i(g) - u_i\left(g - ig(i)\right), 0\right\}\right] = \min_{i:g(i)\neq\varnothing}\left[\max\left\{u_i(g) - u_i(\varnothing), 0\right\}\right]. \tag{4}$$

For the logit choice rule, $c_L(g)$ is therefore the minimum of the quantities in (3) and (4).

Example 4.1. Suppose that $M = \{m_1, m_2, m_3\}$, $W = \{w_1, w_2, w_3\}$, and that the matrix giving their payoffs from a given match is shown below. For example, the top left cell tells us that m_1 gets a payoff of 10 from being matched with w_1.[15] The payoffs from being single are zero for all men and women. Let the perturbed dynamic be the logit choice rule.

	w_1	w_2	w_3
m_1	10,1	5,5	1,10
m_2	1,10	10,1	5,5
m_3	5,5	1,10	10,1

There are three stable matchings as below.

$$g_1 = \{m_1w_1, m_2w_2, m_3w_3\}, \quad g_2 = \{m_1w_2, m_2w_3, m_3w_1\}, \quad g_3 = \{m_1w_3, m_2w_1, m_3w_2\}.$$

Note that g_1 is man-optimal and g_3 is woman-optimal. Also note that $c_L(g) = 1$ for $g \in \{g_1, g_3\}$. For example, one of the least cost deviations from g_1 is w_1 becoming single, which costs 1. Let g' denote the resulting matching. The cost of this deviation is:

$$c_L(g_1) = c\left(g_1, g'\right) = u_{w_1}(g_1) - u_{w_1}\left(g'\right) = 1 - 0 = 1.$$

Followed by $\{m_1w_2\}$, $\{m_2w_3\}$, $\{m_3w_1\}$ matching sequentially, the dynamic will reach g_2.

Moreover, $c_L(g_2) = 4$. One of the least cost deviations from g_2 is for m_1 and w_1 to partner, causing the payoff of w_1 to decrease by 4. These values for $c_L(.)$ imply that $OS = \{g_2\}$. So $SS = \{g_2\}$, the unique stochastically stable matching is g_2.

Under the logit choice rule, OS corresponds to a non-transferable utility version of the *least-core* [25].

Definition 4.2. For $A \subseteq N$, let $G(A, g)$ be the set of matchings $g' \neq g$ such that $ij \in g'$ for all $i, j \notin A$, $ij \in g$, and $ij \notin g'$ for all $i \notin A$, $ij \notin g$. Then the excess of A at g is defined as

[15] This example is a cardinal utility version of Example 1 of Gale and Shapley [16].

$$e(A,g) := \max_{g' \in G(A,g)} \sum_{i \in A} \min\{0, u_i(g') - u_i(g)\},$$

and the least-core is

$$\mathfrak{LC} := \arg\min_{g \in \mathfrak{C}} \max_{A\colon G(A,g) \neq \varnothing} e(A,g).$$

Note that in contrast to the definition of excess in Maschler et al. [25], we do not allow players' gains to enter the calculation. Within the core, excess is a measure of the amount by which a constraint is satisfied, and in a non-transferable utility setup this is unaffected by potential gains in payoff. In marriage problems, the maximum excess can be found by analyzing A such that $|A| \leq 2$. The following proposition follows immediately.

Proposition 4.3. *Under the logit choice rule, $OS = \mathfrak{LC}$.*

We can further characterize properties of matchings in OS for generic payoffs under logit and similar choice rules. Suppose a set of payoff vectors which satisfy, for all $i, j, i', j' \in N$, $g, g' \in G$,

$$u_i(g \pm ij) - u_i(g) = u_{i'}(g' \pm i'j') - u_{i'}(g') \quad \Rightarrow \quad i = i',\ j = j',\ g(i) = g'(i'),$$

where $g \pm xy$ is $g + xy$ if $xy \notin g$ and $g - xy$ otherwise. The payoffs are generic in the sense that the complement of the closure of such a set has Lebesgue measure zero in $\mathbb{R}^{|G| \times |N|}$ satisfying payoff assumptions in Section 2.1.

Remark 4.4. For generic payoffs, under the logit choice rule, the set of pairs of players involved in least cost deviations is a singleton, and is identical across matchings in OS. That is, $N_{\mathrm{L}}(g) = N_{\mathrm{L}}(g')$ for all $g, g' \in OS$. This implies that $g(i) = g'(i)$ and $g(j) = g'(j)$ for $(i, j) \in N_{\mathrm{L}}(g)$ for all $g, g' \in OS$.

4.2. Comparison of alternative rules

By means of an example, we now consider three popular choice rules successively. In doing so we show differences and subtleties, highlighting the difference between the current work and previous work which considers only uniform errors [19,21].

Suppose that $M = \{m_1, m_2, m_3\}$, $W = \{w_1, w_2, w_3\}$, and that the matrix giving players' payoffs from a given match is given below. Payoffs from remaining single are assumed to be zero. The stable matchings are $g_W = \{m_1w_2, m_2w_1, m_3w_3\}$ and $g_M = \{m_1w_1, m_2w_2, m_3w_3\}$. g_W is the woman optimal matching and g_M the man optimal matching.

	w_1	w_2	w_3
m_1	10,5	5,10	6,1
m_2	5,10	10,5	2,10
m_3	3,3	2,3	5,5

4.2.1. Uniform mistakes

Consider a move from g_W to g_M. Earlier work on uniform mistakes would consider the simplest sequence of transitions from g_W to g_M. For example, Step 2 in the proof of Theorem 2 of Klaus et al. [21] could be interpreted as follows. Since $w_2 = g_W(m_1) \neq g_M(m_1) = w_1$, the pair (m_1, w_1) is chosen to match. This is a mistake as w_1 loses payoff. The resulting matching is $g_2 = \{m_1w_1, m_2, w_2, m_3w_3\}$. Note that $m(g_2, g_M) > m(g_W, g_M)$. That is, the new matching is closer to g_M than was the original matching. From g_2, the unperturbed dynamic can costlessly reach g_M. One mistake which is carefully chosen to increase $m(\cdot, g_M)$ can move the process from g_W to g_M. A similar trick is used to move from g_M to g_W, and as every mistake has the same cost under uniform errors, $OS = \{g_W, g_M\}$. Further, as discussed in Remark 3.5, $SS = \{g_W, g_M\}$. Note that when both players are mistaken in forming a match, it does not matter whether this counts as one or two errors: as we have just seen, on the required paths, it is never the case that both players who form a match are mistaken in doing so.

4.2.2. Logit choice rule

When employing payoff-dependent (or otherwise differing) mistakes, we are treading into more tricky territory. Theorem 3.4 shows we can, indeed must, restrict our attention to particular deviations which incur the minimum cost. The least cost deviation from g_W under the logit choice rule has $\{m_2w_3\}$ forming a link. Denote the resulting matching $g_1 = \{m_1w_2, m_2w_3, w_1, m_3\}$. m_2 loses 3 units of payoff from the match, and w_3 does not make a loss, so $c(g_W, g_1) = 3 + 0 = 3$. Observe that g_1 is more distant from g_M than was the original matching: $m(g_1, g_M) < m(g_W, g_M)$. What Lemma 3.3 shows is that even from g_1, the process can costlessly reach some state closer to g_M than g_W is. Fig. 1 earlier in the paper illustrates the contrast between the paths used when considering uniform errors and the paths which must be used to prove Theorem 3.4. In the example under consideration, following the move to g_1, it can be the case that $\{m_1w_1\}$, $\{m_2w_2\}$, $\{m_3w_3\}$ match sequentially, reaching g_M. These transitions have zero cost. So, a single least cost deviation suffices to move the process from g_W to g_M. The least cost deviation from g_M involves m_3 and w_1 matching. They each make a payoff loss of 2 by forming this match, so the cost of the mistake is $2 + 2 = 4$. We conclude that $OS = \{g_M\}$, so g_M is the unique stochastically stable matching.

4.2.3. Probit choice rule

The reader may conjecture that any mistake model in which the cost of mistakes is increasing in payoff loss will always give the same OS as the logit choice rule. Such a conjecture is false, as we now show. The conjecture is true if we restrict attention to error models in which every mistake only involves loss of payoff by a single player. However, the possibility of two players erring at the same time complicates matters. To see this, consider the probit choice rule. Specifically, consider a perturbed dynamic similar to that in Section 4.1, differing in that players decide whether to accept new matchings or leave existing matchings according to probit instead of logit choice. Dokumaci and Sandholm [11] show that under probit choice, mistake costs are proportional to the square of payoff loss. However, the combined cost of two players making a mistake will still be additive. The example of this section has been constructed so that least cost deviations remain the same under logit and probit. Calculating, we see that $c_L(g_W) = 3^2 + 0^2 = 9$ and $c_L(g_M) = 2^2 + 2^2 = 8$. That is, under probit choice, $OS = \{g_W\}$, so g_W is the unique stochastically stable matching. Note that compared to logit, the convexity of costs in probit is friendly towards mistakes by multiple players.[16]

More generally, OS under the probit choice rule coincides with a variant $\mathfrak{LC}_{\text{pro}}$ of the least-core (Definition 4.2) with excess renormed as

$$e_{\text{pro}}(A,g) := \max_{g' \in G(A,g)} -\left(\sum_{i\in A}\left[\min\left\{0, u_i\left(g'\right) - u_i(g)\right\}\right]^2\right)^{\frac{1}{2}}.$$

Recalling that $-e(A,g)$ is a minimum sum of potential payoff losses of members of A, $-e_{\text{pro}}(A,g)$ can be interpreted as the minimum Euclidean length of vectors of potential payoff losses. Hence the difference between logit and probit when it comes to determining OS is equivalent to the difference between using the taxicab[17] and Euclidean norms to assess the size of a vector of payoff losses.

4.2.4. Non-genericity of predictions of uniform mistake models

Consider a choice rule for which individual mistake costs are proportional to payoff loss to the power of $a \geq 0$. In the special cases of $a = 0, 1, 2$, we have the uniform, logit and probit cases respectively. Considering our example above, if $b = \log 2/(\log 2 - \log 3)$, then $OS = \{g_M\}$ for $0 < a < b$, and $OS = \{g_W\}$ for $a > b$. If $a = 0$ or $a = b$, then $OS = \{g_W, g_M\}$. Results for the uniform case are non-generic in this class of models.

4.2.5. $OS \neq SS$

Least cost deviations are always enough to move from stable matchings outside OS towards stable matchings in OS, but the same is not necessarily true when moving between matchings in OS. This fact implies that there exist cases in which $OS \neq SS$.

Let w_3 be a slightly eccentric individual who is particularly prone to leaving her partner.[18] Let the cost of any mistake in which w_3 leaves her partner be 1. Let the cost of all other mistakes be 2. Then a least cost deviation from both g_W and g_M is for w_3 to leave m_3 and become single at a cost of 1. Therefore $OS = \{g_W, g_M\}$. Starting from g_W, following this least cost deviation, $\{m_1w_3\}$, $\{m_2w_2\}$, $\{m_1w_1\}$, $\{m_3w_3\}$ can match sequentially at zero cost, reaching g_M. However, starting from g_M, following the least cost deviation, the only costless transition is for $\{m_3, w_3\}$ to rematch, returning to g_M. To move from g_M to g_W a more costly error is required. Hence $SS = \{g_M\} \neq OS$.

5. Many-to-one matching problems

We extend our analysis to many-to-one matching problems, also known as college admissions problems. The difference from one-to-one matching problems is that each player from one population, the colleges, may be matched with more than one player from the other population, the students. Each student is matched with at most one college.

There are two sets, $\mathbf{K} = \{K_1, \ldots, K_l\}$ and $S = \{s_1, \ldots, s_m\}$, of colleges and students respectively. There is positive integer q_K, called the quota, of college K which indicates the maximum

[16] So when it comes to making mistakes, the adage 'misery loves company' is more true under probit choice than under logit.

[17] The taxicab norm sums the elements of a vector.

[18] We assume such an individual so as to use the same example and make the discussion simpler. Examples for which $OS \neq SS$ can be shown for logit and probit dynamics.

number of positions college K may fill. That is, $|g(K)| \leq q_K$. All q_K positions of college K are identical. The set of matchings in the college admissions problem is:

$$G_{CA} = \big\{g \in \mathcal{G} : (\forall ij \in g, i \in S \Leftrightarrow j \in \mathbf{K}),\ \big(\forall i \in S, \big|g(i)\big| \leq 1\big), \\ \big(\forall K_j \in \mathbf{K}, \big|g(K_j)\big| \leq q_{K_j}\big)\big\}.$$

The preferences of college K are determined by the subset of students to which K is matched. That is, although $g(K)$ can now be of size greater than one, it is still the case that $g(K) = g'(K)$ implies that $u_K(g) = u_K(g')$. Preferences over subsets of students are still assumed to be strict: $g(K) \neq g'(K) \Leftrightarrow u_K(g) \neq u_K(g')$.

Definition 5.1. A matching $g \in G_{CA}$ is in the core, denoted $g \in \mathfrak{C}'$, if $\nexists A \subseteq N$, $g' \in G_{CA}$ such that:

(i) $i \notin A,\ j \notin A,\ ij \in g \Rightarrow ij \in g'$,
(ii) $ij \notin g,\ ij \in g' \Rightarrow i \in A,\ j \in A$,
(iii) $i \in A \Rightarrow u_i(g') > u_i(g)$.

We restrict our attention to responsive preferences [32]. If a college has responsive preferences, then its preferences over any two students s_i, s_j are independent of the other students to whom it is matched. That is, if a college, K, prefers s_i to s_j, and T, $|T| < q_K$, is some subset of students which includes neither s_i nor s_j, then the college prefers $T \cup s_i$ to $T \cup s_j$. We assume that all colleges have responsive preferences.

Definition 5.2. The preferences of college $K \in \mathbf{K}$ over sets of students are responsive if they satisfy the following conditions.

(I) If $g(K) = g'(K) \cup \{s_i\} \setminus \{s_j\}$, $s_i \notin g'(K)$, $s_j \in g'(K)$, then $u_K(\{Ks_i\}) > u_K(\{Ks_j\}) \Leftrightarrow u_K(g) > u_K(g')$.
(II) If $g(K) = g'(K) \cup \{s_i\}$, $s_i \notin g'(K)$, then $u_K(\{Ks_i\}) > u_K(\varnothing) \Leftrightarrow u_K(g) > u_K(g')$.

Following Chapter 5 of Roth and Sotomayor [34], we consider a related marriage problem, in which each college K is broken into q_K positions of itself: $k_1, \ldots, k_{q_K}$, each of which has a quota of one. In the related market, the players are students and college positions each of which has a quota of one. The college positions are assumed to have the same preferences over the individual students as their original college. Students are assumed to be indifferent between positions in the same college.

The results of Roth and Sotomayor [34][19] then imply that the core (Definition 5.1) of the college admissions problem is related to the set of stable matchings (Definition 2.1) of the associated marriage problem in the following way.

Remark 5.3. Let $g' \in G_{CA}$ be a network for a college admissions problem and $g \in G$ be a network for the associated marriage problem. If, for all $K \in \mathbf{K}$ with associated positions $\{k_1, \ldots, k_{q_K}\}$, we have $g'(K) = \bigcup_{1 \leq i \leq q_K} g(k_i)$, then $g' \in \mathfrak{C}'$ if and only if $g \in \mathfrak{C}$.

[19] Lemma 5.6 and Proposition 5.36. Note that unlike Proposition 5.36 of the cited paper, our dominance criterion in Definition 5.1 does not have to be weak, as we, like Jackson and Watts [19], allow colleges within a deviating coalition to remain matched to students outside of the coalition.

Henceforth, with a slight abuse of notation, we let K denote the set of positions in college K, i.e. $K = \{k_1, \dots, k_{q_K}\}$, $g(K) = \bigcup_{1 \le i \le q_K} g(k_i)$.

Definition 5.4. Define the set of matchings equivalent to $g \in G$ as:

$$Eq(g) = \{g' \in G : g'(K) = g(K)\ \forall K \in \mathbf{K}\}.$$

In words, $Eq(g)$ is the set of matchings in which students are matched to the same colleges as they are in matching g, i.e. matchings in $Eq(g)$ are identical in the original college admissions problem.

Take any unstable matching $g \notin \mathfrak{C}$, and a target stable matching $g' \in \mathfrak{C}$. The following lemma, which is important to the results of this section, shows that, starting from g, the unperturbed dynamic can move to some matching g_T which is strictly closer to g' than g is. This lemma extends the implications of Lemma 5 of Klaus et al. [21] to many-to-one matching problems. First, define a similarity function for the many-to-one matching problem:

$$\bar{m}(g, g') := \max_{\hat{g} \in Eq(g')} m(g, \hat{g}). \tag{5}$$

Note that $\bar{m}(g, g') \ge m(g, g')$. Also note that $m(.,.) \equiv \bar{m}(.,.)$ for one-to-one matching problems.

Lemma 5.5. *Let $g \notin \mathfrak{C}$, $g' \in \mathfrak{C}$. Then, $\exists T \in \mathbb{N}_+$, $g_T \in G$, such that $P_0^T(g, g_T) > 0$ and $\bar{m}(g_T, g') > \bar{m}(g, g')$.*

The proof is left to Appendix B, and makes use of the fact that, given g and g', there is no student matched to different positions of the same college under g and the g^* which solves the maximization in (5).

Lemma 5.6. *Let $g \notin \mathfrak{C}$, $g' \in \mathfrak{C}$. Let*

$$g^* \in \operatorname*{argmax}_{\hat{g} \in Eq(g')} m(g, \hat{g}).$$

For all $i \in S$, $g(i) \in K$, $g^*(i) \in K \Rightarrow g(i) = g^*(i)$.

Proof. Assume $i \in S$, $g(i) \in K$, $g^*(i) \in K$, $g(i) \neq g^*(i)$. Let $g^{**} = g^* + ig(i) + g^*(i)g^*(g(i))$. Then $g^{**} \in Eq(g^*) = Eq(g')$ and $m(g, g^{**}) \ge m(g, g^*) + 2$, contradicting the definition of g^*. □

The proof of Lemma 5.5 relies on the construction of closed cycles of players who have strict preferences between g and g^*. Lemma 5.6 ensures that players who have the same partner in g and g^*, and who are therefore indifferent between the two matchings, form separate cycles of size two.

Lemma 5.5 directly implies the following corollary. It is similar to Roth and Vande Vate [35], except that we have not assumed students to have strict preferences over the positions within colleges.[20,21]

[20] See Chapter 5 of Roth and Sotomayor [34] for a way to construct strict preferences in such problems.

[21] As students only ever match with a single college, their preferences are substitutable. Therefore, the many-to-many paths to stability result of Kojima and Ünver [23] also implies this corollary.

Corollary 5.7 *(Random paths to stability). Suppose a college admissions problem, its related marriage problem, and the unperturbed dynamic of Section 2.2. For any $g \notin \mathfrak{C}$, there exists $T \in \mathbb{N}_+$, $g^* \in \mathfrak{C}$, such that $P_0^T(g, g^*) > 0$.*

Under the unperturbed dynamic, a player will only change partner if such a change leads to a strict increase in utility. Therefore, it is never the case that positions in the same college compete for students. This is realistic, as in the original problem, college K does not distinguish between a student filling position $k_i \in K$ or $k_j \in K$.[22] We now impose a similar restriction on the perturbed dynamic.

Assumption 4. Let $\nu(g) := \{(i, k_j) : k_j \in K \in \mathbf{K}, i \neq g(k_j), i \in g(K)\}$. In Condition (v) of Assumption 1, replace the set of possible deviating pairs μ with $\mu \setminus \nu(g^t)$.

Note that the logit dynamic under Assumption 4 is still irreducible. For any pair $ik_j \in g$, we have $(i, k_j) \notin \nu(g)$, so with positive probability i and k_j will separate. In such a manner, the empty network can be attained within $|S|$ periods. Furthermore, if $g(i) = g(k_j) = \varnothing$, we have $(i, k_j) \notin \nu(g)$, so starting from the empty network, any network in G can be attained within a further $|S|$ periods.

We make a natural symmetry assumption on the dynamic regarding the behavior of positions of a college. We assume that the cost of transitions is unaffected by the labeling of the positions of any given college.

Assumption 5. If $\tilde{g} \in Eq(g)$; $k_1, k_2 \in K_i : g(k_1) = \tilde{g}(k_2)$; $s \in S : g(s), \tilde{g}(s) \in K_j \in \mathbf{K}$ or $g(s) = \tilde{g}(s) = \varnothing$; then:

(i) $c(g, g + k_1 s) = c(\tilde{g}, \tilde{g} + k_2 s)$,
(ii) If $g(k_1) \neq \varnothing$, then $c(g, g - k_1 g(k_1)) = c(\tilde{g}, \tilde{g} - k_2 \tilde{g}(k_2))$, and
(iii) If $g(s) \neq \varnothing$, then $c(g, g - sg(s)) = c(\tilde{g}, \tilde{g} - s\tilde{g}(s))$.

Note that the logit choice rule satisfies Assumption 5.

Define $c_{\mathrm{L}}(g)$ and OS as in the one-to-one matching problem. Using Lemma 5.5, under Assumptions 4 and 5, a many-to-one version of Lemma 3.3 can be proved. Then, we have the following theorem. See Appendices A and B for proofs.

Theorem 5.8. *Under Assumptions 4 and 5, $SS \subseteq OS$.*

Example 5.9 *(The logit choice rule and Assumption 4).* Assumption 4 implies that pairs in $\nu(g)$ may not deviate if the process is at g. That is, (i, k_j), $k_j \in K$, may not deviate if i already occupies a position $k_l \neq k_j$ at college K. For the logit dynamic, expressions for $c_{\mathrm{L}}(g)$ will be as in expressions (3) and (4), but with the minimum in expression (3) taken over $ik_j \notin g \cup \{ik_j : (i, k_j) \in \nu(g)\}$.

We conclude this section with three examples. Example 5.10 is an application of Theorem 5.8 with a note emphasizing the role of Assumption 4. Examples 5.11 and 5.12 demonstrate that

[22] There may exist cases in which different departments of a college compete for students. In such cases, we let K and K' be such that $K \neq K'$ represent different departments.

extending our result beyond responsive preferences, for example to settings in which colleges have a preference for homogeneity in their student bodies, is not straightforward.

Example 5.10. Let $S = \{s_1, s_2, s_3\}$, $\mathbf{K} = \{K, K'\}$, $K = \{k_1, k_2\}$ and $K' = \{k_3\}$. Assume that a college's utility is additive over the utility it obtains from each student, and that the perturbed dynamic is the logit choice rule. Preferences are given by the following matrix. The payoff from remaining unmatched is assumed to be zero.

	k_1	k_2	k_3
s_1	10,10	10,10	5,10
s_2	10,8	10,8	6,8
s_3	5,9	5,9	10,5

Observe that the set of stable matchings is $\mathfrak{C} = \{g_1, g_2, g_3, g_4\}$ where

$$g_1 = \{(s_1, k_1), (s_2, k_2), (s_3, k_3)\}, \qquad g_2 = \{(s_1, k_2), (s_2, k_1), (s_3, k_3)\},$$
$$g_3 = \{(s_1, k_1), (s_2, k_3), (s_3, k_2)\}, \qquad g_4 = \{(s_1, k_2), (s_2, k_3), (s_3, k_1)\}.$$

The first two matchings are equivalent, $g_2 \in Eq(g_1)$. s_1 and s_2 are matched to K in both matchings. Similarly, $g_4 \in Eq(g_3)$.

Suppose that the current network is g_1. In the absence of Assumption 4, a deviation by (s_1, k_2) to $g_1 + s_1k_2$ could occur with cost zero. Subsequently, (s_2, k_3) and (s_3, k_1) could form partnerships, and the process could reach g_4 without any additional cost. So $C(g_1, g_4)$ would equal zero. Similarly, we can cycle between all of the matchings in $\mathfrak{C}$.

Under Assumption 4 (s_1, k_2) will never be selected as a revising pair when the current state is g_1. The least cost deviation from g_1 is $\mathrm{L}(g_1) = \{g_1 + s_2k_3\}$ with cost $c_{\mathrm{L}}(g_1) = 4$. Also, $c_{\mathrm{L}}(g_2) = 4$, $c_{\mathrm{L}}(g_3) = 1$, $c_{\mathrm{L}}(g_4) = 1$. $OS = \{g_1, g_2\}$. Since g_1 and g_2 are equivalent, the unique stochastically stable matching is that K and K' are matched to $\{s_1, s_2\}$ and s_3 respectively.

Example 5.11 *(Non-responsive preferences).* Consider a college admissions problem with $S = \{s_1, s_2, s_3\}$, $\mathbf{K} = \{K\}$ and $q_K = 3$. Let $u_x(K) = 10$ and $u_x(\varnothing) = 0$ for all $x \in S$. Also, for $X \subseteq S$, let

$$u_K(X) = \begin{cases} 5 & \text{if } |X| = 3, \\ 4 & \text{if } |X| = 2, \\ -2 & \text{if } |X| = 1, \\ 0 & \text{if } X = \varnothing. \end{cases} \tag{6}$$

In words, each student prefers being in K to being out. College K prefers to have at least two students to none, but prefers none to having one student only. Let the perturbed dynamic be the logit choice rule.

There are two stable matchings: one where K accepts all students and another where none are accepted. The former is uniquely stochastically stable, while the latter is most robust to one-shot deviation. To see this, observe that one costly deviation, which costs 2, is enough to move from none accepted to all accepted. While at least two costly deviations, which cost 7, are required in the opposite move. Lemma 5.5 does not apply here and hence Theorem 5.8 does not hold.

Note that in the absence of responsive preferences, Remark 5.3 does not hold. In Example 5.11, we just saw that although the core contains a unique matching in which every student matches with the college, there is an additional stable matching in which no students match with the college. It may be argued that, in the absence of the equivalence of Remark 5.3, a richer process of strategic updating should be used. The following example allows groups of players to rematch amongst themselves each period, with no limitations on the size of such a group. In the presence of two types of student, the colleges prefer homogeneous student bodies. These preferences satisfy substitutability,[23] but violate responsiveness, and Theorem 5.8 still fails to hold.

Example 5.12 *(A desire for homogeneity).* Consider a college admissions problem with three students $S = \{s_{y1}, s_{y2}, s_{p3}\}$, and two colleges $\mathbf{K} = \{K_y, K_p\}$, $q_{K_y} = q_{K_p} = 2$. The students are either Yellow students (s_{y1}, s_{y2}), or Pink students (s_{p3}). Yellow students prefer college K_y, and Pink students prefer college K_p. Let $u_{s_{y1}}(K_y) = u_{s_{y2}}(K_y) = u_{s_{p3}}(K_p) = 20$, $u_{s_{y1}}(K_p) = u_{s_{y2}}(K_p) = u_{s_{p3}}(K_y) = 10$, $u_x(\varnothing) = 0$ for all $x \in S$. Let the utilities of the colleges, which do not satisfy responsiveness, but do satisfy substitutability, be given by the following table.

$X \subseteq S$	$\{s_{y1}, s_{y2}\}$	$\{s_{y1}\}$	$\{s_{y2}\}$	$\{s_{p3}\}$	$\varnothing$
$u_{K_p}(X)$	5	4	3	2	0
$u_{K_y}(X)$	5	2	1	8	0

The utilities of both colleges from the heterogeneous sets $\{s_{y1}, s_{p3}\}$, $\{s_{y2}, s_{p3}\}$ are assumed to be negative. Both colleges prefer a homogeneous student population to no students, and prefer no students to a heterogeneous student population. The colleges' preferences are opposed to those of the students. College K_p prefers Yellow students to Pink. College K_y prefers Pink to Yellow. The core of the problem contains two matchings, $g^\dagger$ and $g^\ddagger$. In $g^\dagger$, students attend their preferred colleges. In $g^\ddagger$, they do not. That is,

$$g^\dagger(K_y) = \{s_{y1}, s_{y2}\} = g^\ddagger(K_p), \qquad g^\ddagger(K_y) = \{s_{p3}\} = g^\dagger(K_p).$$

Let the process of rematching be as follows. Each period some subset of players $A \subseteq N$ is chosen. Let $g^t = g$ be the current matching. A conjectured rematching g' which satisfies (i) and (ii) of Definition 5.1 is chosen at random and accepted by each member of A with probabilities given by the logit choice rule. If any member of A rejects the rematching, then $g^{t+1} = g^t = g$. If every member of A accepts the rematching, then $g^{t+1} = g'$.

From $g^\dagger$, the least cost deviation involves college K_p expelling student s_{p3} at cost $c_L(g^\dagger) = 2 - 0 = 2$. Following such a deviation, a zero cost path to $g^\ddagger$ exists: K_y simultaneously expels s_{y1} and s_{y2} while accepting s_{p3}, following which K_p accepts s_{y1} and s_{y2}.

From $g^\ddagger$, the least cost deviation involves college K_p expelling student s_{y2} at cost $c_L(g^\ddagger) = 5 - 4 = 1$. However, following such a deviation, there is no zero cost path to $g^\dagger$. Reaching $g^\dagger$ requires at least one further costly deviation, such as college K_p simultaneously expelling s_{y1} and accepting s_{p3}. This has an additional cost of 2, making a total cost of 3.

So we have that $OS = \{g^\dagger\}$, yet $SS = \{g^\ddagger\}$. Theorem 5.8 does not hold.

[23] In the context of cardinal preferences, college K's preferences are substitutable if $S_1^* \cap S_2 \subseteq S_2^*$ for all $S_2 \subseteq S_1 \subseteq S$, where $S_i^* = \text{argmax}_{S' \subseteq S_i : |S'| \le q_K} u_K(S')$. That is, a student who is chosen from a larger set of potential students is always chosen from a smaller set.

6. Roommate problems

In the one-sided matching problem, or roommate problem, no distinction is made between men and women. Anyone can partner with anyone. The set of networks of interest is broadened to:

$$G_R = \{g \in \mathcal{G} : (\forall i \in N, |g(i)| \leq 1)\}.$$

Gusfield and Irving [18] show that two key properties of marriage problems extend to all roommate problems with strict preferences over partners. Firstly, the set of unmatched players is the same at every stable matching. Secondly, if $g, g' \in \mathfrak{C}$, i prefers g to g', $g(i) = j$, $g'(i) = k \neq j$, then both j and k prefer g' to g. These properties are exactly those used in our results of Section 3. Furthermore, Diamantoudi et al. [10] show that if $\mathfrak{C}$ is nonempty, then there exists a sequence of mutually beneficial blockings ending in $\mathfrak{C}$. In the context of this paper, this means that nonempty $\mathfrak{C}$ implies that all recurrent classes of the unperturbed Markov process lie in $\mathfrak{C}$. There are no absorbing cycles. Assuming nonemptiness of $\mathfrak{C}$, Lemmas 3.1, 3.3 still hold. It follows that:

Theorem 6.1. *If $\mathfrak{C} \neq \emptyset$, then $SS \subseteq OS$.*

Thus our main result does not rely on two-sidedness of the matching market.

7. Conclusion

This paper has shown that in marriage problems, roommate problems and college admissions problems, all stochastically stable matchings are in the class of matchings which are most robust to one-shot deviation. There are two significant implications of this from a market design perspective. Firstly, a desired matching may not be stochastically stable, so even if implemented in the short run, in a world in which people make the occasional mistake, it would be rarely observed in the long run. Secondly, making a desired matching more robust to one-shot deviation than any other matching will suffice to make it uniquely stochastically stable. The main results, which link stochastic stability to a local property of the individual matchings, are derived from the structure of stable matchings and from the unperturbed blocking dynamic. The class of unperturbed blocking dynamics we use is common in the paths to stability literature. Further attempts to extend our results to, for example, hedonic games or many-to-one matchings with preferences beyond responsiveness, are left for future work.

Acknowledgments

The authors would like to thank Michihiro Kandori, Bettina Klaus, Bettina Klose, Heinrich Nax, Hiroyuki Ozaki, Bary Pradelski, William Sandholm, Jun Xiao and Peyton Young for detailed comments, as well as seminar audiences at Birkbeck, Queen Mary and Royal Holloway, University of London, Kanagawa University (JEA Fall 2013), Keio University, Tokyo Institute of Technology (Game Theory Workshop 2014), University of Oxford, University of Adelaide, Monash University, University of Essex and University of York. The authors also thank the Associate Editor and two anonymous referees for helpful insights and questions.

Appendix A. One-to-one matchings

In this section, we prove Lemma 3.3 and Theorem 3.4. These results are implied by our more general results for the college admissions problem (Lemma B.3 and Theorem 5.8), but proofs for the simpler case are included to ease clarity of exposition. Conveniently, they also serve as proofs for Theorem 6.1 (roommate markets – see discussion in text).

The proof uses Lemma 3.1, Lemma 5 of Klaus et al. [21], and the following two properties of stable matchings: (i) the set of unmatched players is the same at every stable matching, so if i is matched at $g \in \mathfrak{C}$, then i being single at g' implies g' is unstable; (ii) if $g, g' \in \mathfrak{C}$, i prefers g to g', $g(i) = j$, $g'(i) = k \neq j$, then both j and k prefer g' to g (see [18,34]).

Proof of Lemma 3.3.
Step 1: We show that for $g_1 \in \mathrm{L}(g)$, there exists $T \in \mathbb{N}_+$, $g_T \in G$, such that $P_0^T(g_1, g_T) > 0$, $m(g^*, g_T) \geq m(g^*, g)$, and $g_T \notin \mathfrak{C}$.

Suppose that $g_1 = g - ig(i) \in \mathrm{L}(g)$. By Lemma 3.1, $g \notin OS$ implies $g(i) \neq g^*(i)$, so $m(g^*, g_1) = m(g^*, g)$. As i is single at g_1, g_1 is unstable.

Next, suppose that $g_1 = g + ij \in \mathrm{L}(g)$. Lemma 3.1 implies that $g(i) \neq g^*(i)$ and/or $g(j) \neq g^*(j)$.

Case I: $ij \in g^*$.

Note that $m(g^*, g_1) > m(g^*, g)$. $ij \in g^*$ implies $g(i) \neq \emptyset$. $g + ij \in \mathrm{L}(g)$ implies $g(i) \neq j$. Therefore $g(i)$ is single at g_1 so g_1 is unstable.

If $ij \notin g^*$, then we have two cases:

Case IIa: $g(i) \neq g^*(i)$ and $g(j) \neq g^*(j)$.

Note that $m(g^*, g_1) = m(g^*, g)$. As $g(i)$ is single at g_1, g_1 is unstable.

Case IIb: $g(i) = g^*(i)$ and $g(j) \neq g^*(j)$.

Note that $m(g^*, g_1) \geq m(g^*, g) - 2$. First, suppose that i prefers g to g_1. If $g(i) \neq \emptyset$, let i and $g(i)$ get matched. If $g(i) = \emptyset$, let i leave j to be single. Let g_2 denote the resulting network. $g(j) \neq g^*(j)$ implies that $m(g^*, g_2) = m(g^*, g)$. As $g(j)$ is single at g_2, g_2 is unstable.

Next, suppose that i prefers g_1 to g (and therefore g^*). Therefore, as $g, g^* \in \mathfrak{C}$, it must be that j prefers g and g^* to g_1.

If j prefers g^* to g, then $g(j)$ prefers g to g^*. This implies that $g^*(g(j))$ prefers g^* to g. Let $g^*(g(j))$ and $g(j)$ get matched. Let g_3 denote the resulting network. Note that $m(g^*, g_3) \geq m(g^*, g)$. $g(g^*(g(j)))$ is single at g_3, so g_3 is unstable.

If j prefers g to g^*, then $g^*(j)$ prefers g^* to g. Let j and $g^*(j)$ get matched. Let g_4 denote the resulting network. Note that $m(g^*, g_4) \geq m(g^*, g)$. As $g(j)$ is single at g_4, g_4 is unstable.

Step 2: Given the g_T from Step 1, Lemma 5 of Klaus et al. [21] implies the existence of $T_1 \in \mathbb{N}_+$, $g_{T_1} \in G$, such that $P_0^{T_1}(g_T, g_{T_1}) > 0$, $m(g^*, g_{T_1}) > m(g^*, g_T)$. If $g_{T_1} \in \mathfrak{C}$ then set $g' = g_{T_1}$. Otherwise, apply Lemma 5 of Klaus et al. [21] again to attain g_{T_2} such that $m(g^*, g_{T_2}) > m(g^*, g_{T_1})$. As $m(g^*, .)$ is bounded above by $|N|$, we must eventually reach some $g_{T_m} \in \mathfrak{C}$ such that $m(g^*, g_{T_m}) > m(g^*, g)$. □

Proof of Theorem 3.4. If $g \in SS$, then $g \in \mathfrak{C}$ and there exists a minimal cost spanning tree rooted at g. Denote the cost of this tree by $cost(g)$. Assume $g \notin OS$. Choose $g^* \in OS$. Construct a path of edges $(g = g_1, \dots, g_L)$ such that $g_i \in \mathfrak{C}$, $g_i \notin OS$ for $i = 1, \dots, L-1$, and $g_L \in OS$. The path is constructed as follows. For each g_i, $i = 1, \dots, L-1$, Lemma 3.3 implies:

$$\exists g_{i+1} \in \mathfrak{C}: \quad m\big(g^*, g_{i+1}\big) > m\big(g^*, g_i\big) \quad \text{and} \quad C(g_i, g_{i+1}) = c_{\mathrm{L}}(g_i).$$

This is repeated until we reach some $g_L \in OS$. Add these edges to the conjectured minimal cost spanning tree, replacing the existing edges exiting $g_2, \dots, g_{L-1}$. Remove the edge exiting g_L. Denote the cost of the new tree by $cost(g_L)$. Then:

$$cost(g_L) \le cost(g) + c_{\mathrm{L}}(g) - c_{\mathrm{L}}(g_L) < cost(g).$$

The first inequality follows from the construction of the tree rooted at g_L; the second inequality holds as $g \notin OS$ implies $c_{\mathrm{L}}(g) < c_{\mathrm{L}}(g_L)$. So, the conjectured minimal cost spanning tree can have been no such thing. Contradiction. □

Appendix B. Many-to-one matchings

In this section, we prove Theorem 5.8 via Lemma B.3, an equivalent of Lemma 3.3 for the college admissions problem. Firstly, the proof of Lemma 5.5, one of our key lemmas, is given.

Proof of Lemma 5.5. Let

$$g^* \in \operatorname*{argmax}_{\hat{g} \in Eq(g')} m(g, \hat{g}).$$

If there exists $i \in N$ such that $g(i) \neq \varnothing$ and $c(g, g - ig(i)) = 0$ and $g^*(i) = \varnothing$, then let $g_T = g - ig(i)$ and we are done: $\bar{m}(g_T, g') \ge m(g_T, g^*) > m(g, g^*) = \bar{m}(g, g')$.

If there does not exist such an $i \in N$, let each $i \in N$ such that $u_i(\varnothing) > u_i(g)$ leave their partners. Denote the resulting matching g_1. Note that $m(g_1, g^*) = m(g, g^*)$. $g_1 \notin \mathfrak{C}$ as if $g_1 \neq g$, for $i \in S$ such that $g(i) \neq g_1(i) = \varnothing$, $g^*(i) \neq \varnothing$, so i is not single in any stable matching. Note that $g^* \in \operatorname{argmax}_{\hat{g} \in Eq(g')} m(g_1, \hat{g})$. As $g_1 \notin \mathfrak{C}$, $\exists (i, k_j) : c(g_1, g_1 + ik_j) = 0$.

Case I: $\exists (i, k_j)$: $c(g_1, g_1 + ik_j) = 0$ and $ik_j \in g^*$.

Let $g_T = g_1 + ik_j$. Then $\bar{m}(g_T, g') \ge m(g_T, g^*) > m(g_1, g^*) = m(g, g^*) = \bar{m}(g, g')$ and we are done.

Case II: $\forall (i, k_j)$: $c(g_1, g_1 + ik_j) = 0$, $ik_j \notin g^*$.

First, we decompose the player set N into singletons who are unmatched in g_1 and g^*, pairs of players who have the same partner in g_1 and g^*, and cycles defined below. Then, we will construct a path of blockings which increase $\bar{m}(\cdot, g^*)$.

For all $i \in S$: $g_1(i) \in K$, $g^*(i) \in K^*$, $K = K^*$, we have by Lemma 5.6 that $g_1(i) = g^*(i)$.

For all $i \in S$: $g_1(i) \neq g^*(i)$, either $u_i(g_1) > u_i(g^*)$ or $u_i(g^*) > u_i(g_1)$.

Consider i such that $u_i(g_1) > u_i(g^*)$. The arguments when the converse holds are identical. Let $f : N \to N$ be such that $f(j) = g_1(j)$ if $u_j(g_1) > u_j(g^*)$ and $f(j) = g^*(j)$ otherwise.[24] Suppose a sequence $\{i, f(i), f^2(i), f^3(i), \dots\}$ where $f^2(i) = f(f(i))$ and $f^k(\cdot)$ for $k \ge 3$ is defined similarly. Since N is finite, the sequence must repeat and create a cycle. De-

[24] Note that strict preferences and the definition of g^* imply that if $g_1(j) \neq g^*(j)$, then $u_j(g_1) \neq u_j(g^*)$. Furthermore, by $g^* \in \mathfrak{C}$, $g_1(j) = \varnothing \Rightarrow f(j) = g^*(j)$, and by definition of g_1, $g^*(j) = \varnothing \Rightarrow f(j) = g_1(j)$.

note the cycle by a sequence $(n_1, n_2, \ldots, n_m)$, where $n_1 = i$, $n_l = f^{l-1}(n_1)$, and n_m is the last non-repeated element of the cycle. In the sequence, members' preferences alternate between g_1 and g^*, i.e. $g_1(n_j) = n_{j+1}$ if j is odd, and $g^*(n_j) = n_{j+1}$ otherwise.[25] Note that m is even and that $g^*(n_m) = i$ under the assumption that $u_i(g_1) > u_i(g^*)$. Thus, N can be decomposed into singletons, pairs of players and cycles in which players have different partners in g_1 and g^*.

Now, observe that $\nexists (i, k_j)$: $c(g_1, g_2 = g_1 + ik_j) = 0$, $u_i(g_1) \geq u_i(g^*)$ and $u_{k_j}(g_1) \geq u_{k_j}(g^*)$. If there did exist such a (i, k_j), then $u_i(g_2) > u_i(g_1) \geq u_i(g^*)$ and $u_{k_j}(g_2) > u_{k_j}(g_1) \geq u_{k_j}(g^*)$, so (i, k_j) would be a blocking pair for $g^* \in \mathfrak{C}$. So, $u_i(g^*) > u_i(g_1)$ and/or $u_{k_j}(g^*) > u_{k_j}(g_1)$. Without loss of generality, let n_2 be a member of a blocking pair for g_1 such that $u_{n_2}(g^*) > u_{n_2}(g_1)$. Let g_2 be obtained from g_1 by allowing this blocking pair to match. n_2 must be a member of some cycle $(n_1, n_2, \ldots, n_m)$ as defined above, since players who are not in cycles are indifferent between g_1 and g^*. Note that $g_2(g_1(n_2)) = g_2(n_1) = \emptyset$, and that $u_{n_m}(g^*) > u_{n_m}(g_1)$. (n_1, n_m) is a blocking pair for g_2. Let $g_3 = g_2 + n_1 n_m$. $m(g_3, g^*) = m(g_2, g^*) + 2$.

If $g_1(g_2(n_2)) \neq g^*(g_2(n_2))$, then $m(g_2, g^*) = m(g_1, g^*) = m(g, g^*)$, so $\bar{m}(g_3, g') \geq m(g_3, g^*) > m(g, g^*) = \bar{m}(g, g^*)$ and we are done.

If $g_1(g_2(n_2)) = g^*(g_2(n_2))$, then $m(g_2, g^*) \geq m(g_1, g^*) - 2$. If $m \geq 6$, then (n_{m-2}, n_{m-1}) is a blocking pair for g_3 as $g_3(n_{m-1}) = \emptyset$, $u_{n_{m-2}}(g^*) > u_{n_{m-2}}(g_1) = u_{n_{m-2}}(g_3)$. Let $g_4 = g_3 + n_{m-1} n_{m-2}$. Then $\bar{m}(g_4, g') \geq m(g_4, g^*) = m(g_3, g^*) + 2 = m(g_2, g^*) + 4 > m(g_1, g^*) = m(g, g^*) = \bar{m}(g, g')$, and we are done.

For $m = 4$, it cannot be that $u_{n_2}(g_2) > u_{n_2}(g^*)$, or $(n_2, g_2(n_2))$ would be a blocking pair for g^*. If $u_{n_2}(g_2) < u_{n_2}(g^*)$, then (n_2, n_3) is a blocking pair for g_3. Let $g_5 = g_3 + n_2 n_3$. Now $\bar{m}(g_5, g') \geq m(g_5, g^*) = m(g_3, g^*) + 2 = m(g_2, g^*) + 4 > m(g_1, g^*) = m(g, g^*) = \bar{m}(g, g')$, and we are done. If $u_{n_2}(g_2) = u_{n_2}(g^*)$, then $n_3 = g^*(n_2)$, $g_2(n_2)$ are positions in the same college, so when n_2, $g_2(n_2)$ match, $\bar{m}(., g')$ increases by 2. Therefore $\bar{m}(g_2, g') \geq \bar{m}(g_1, g')$. As $\bar{m}(g_3, g') = \bar{m}(g_2, g') + 2$, we have that $\bar{m}(g_3, g') > \bar{m}(g, g')$, and we are done. □

The proof of Lemma 5.5 above implies the following corollary. Over any two stable states $g, g^* \in \mathfrak{C}$ such that $\bar{m}(g, g^*) = m(g, g^*)$, any $i \in N$ such that $g(i) \neq g^*(i)$ has preferences (over g and g^*) in opposition to the preferences of his partners in g and g^*.

Corollary B.1. *Let $g, g' \in \mathfrak{C}$. Let $g^* \in \operatorname{argmax}_{\hat{g} \in Eq(g')} m(g, \hat{g})$. For all $i \in N$ such that $g(i) \neq g^*(i)$, if i prefers g to g^* (g^* to g), then $g(i)$ and $g^*(i)$ prefer g^* to g (g to g^*).*

We now show lemmas analogous to Lemmas 3.1 and 3.3. The next lemma is analogous to Lemma 3.1.

Lemma B.2. *Suppose that $g \in \mathfrak{C}$, $g \notin OS$, $g^* \in OS$ and Assumption 4 holds. Suppose that $(i, k_j) \in N_{\mathrm{L}}(g)$. When it is the case that $g(i) \neq \emptyset$, we shall let K, K^* be such that $g(i) \in K$, $g^*(i) \in K^*$. Let $k_j \in K_j$. Then, we have ($g(i) \neq \emptyset$, $K \neq K^*$) and/or ($\emptyset \neq g(k_j) \notin g^*(K_j)$).*

[25] If $g_1(n_j) = n_{j+1}$, then n_j prefers g_1 to g^*, so n_{j+1} cannot prefer g_1 to g^*, or (n_j, n_{j+1}) would block g^*. If $g^*(n_j) = n_{j+1}$, then n_j prefers g^* to g_1, so n_{j+1} cannot prefer g^* to g_1, or $(n_j, n_{j+1}) \in g^*$ would block g_1.

Proof. Observe that $(\varnothing \neq g(k_j) \notin g^*(K_j))$ if and only if $g(k_j) \neq g^*(k_l)$ for all $k_l \in K_j$. To see this, note that if $g(k_j) = \varnothing$, then $g^*(k_l) = \varnothing$ for some $k_l \in K_j$.[26]

Suppose $(g(i) = \varnothing$ or $g(i) \in K = K^*)$, and $g(k_j) = g^*(k_l)$ for some $k_l \in K_j$. If $g(i) \in K = K_j$, then by Assumption 4, $i = g(k_j)$, so $i = g^*(k_l)$ and $c_{\mathrm{L}}(g^*) \leq c(g^*, g^* - ik_l) = c(g, g - ik_j) = c_{\mathrm{L}}(g)$. If $g(i) = \varnothing$ or $g(i) \in K \neq K_j$, then $c_{\mathrm{L}}(g^*) \leq c(g^*, g^* + ik_l) = c(g, g + ik_j) = c_{\mathrm{L}}(g)$. Therefore $g \in OS$, which contradicts our premise. □

Lemma B.3 *(Getting Closer Lemma II). Suppose the dynamic satisfies Assumptions 4, 5. Let $g' \in OS$. Suppose that $g \in \mathfrak{C}$ and $g \notin OS$. Let $g_1 \in \mathrm{L}(g)$. Then, $\exists g'' \in \mathfrak{C}$, $t \in \mathbb{N}_+$, such that $\bar{m}(g', g'') > \bar{m}(g', g)$ and $P_0^t(g_1, g'') > 0$.*

Proof.
Step 1: We show that for $g_1 \in \mathrm{L}(g)$, there exists $T \in \mathbb{N}_+$, $g_T \in G$, such that $P_0^T(g_1, g_T) > 0$, $\bar{m}(g', g_T) \geq \bar{m}(g', g)$, and $g_T \notin \mathfrak{C}$.

Let g^* satisfy:

$$g^* \in \operatorname*{argmax}_{\hat{g} \in Eq(g')} m(g, \hat{g}) \quad \text{and} \quad g^* \in \operatorname*{argmax}_{\hat{g} \in Eq(g')} m(g_1, \hat{g}).$$

It is possible to choose such a g^* as, by Assumption 4, any student matched to the same college in g and g_1 is matched to the same position of that college. For notation, when it is the case that $g(i) \neq \varnothing$, we shall let K, K^* be such that $g(i) \in K$, $g^*(i) \in K^*$. Let $k_j \in K_j$.

Suppose that $g_1 = g - ig(i) \in \mathrm{L}(g)$, $i \in S$. Under Assumption 5, $g' \in OS$ implies $g^* \in OS$. This, and $g \notin OS$ imply $g(i) \neq g^*(i)$, so $m(g^*, g_1) = m(g^*, g)$, and as $\bar{m}(g', g_1) = m(g^*, g_1)$ and $\bar{m}(g', g) = m(g^*, g)$, we have $\bar{m}(g', g_1) = \bar{m}(g', g)$. As i is single, g_1 is unstable.

Next, suppose that $g_1 = g + ik_j \in \mathrm{L}(g)$. Lemma B.2 implies that $(g(i) \neq \varnothing$ and $K \neq K^*)$ and/or $(\varnothing \neq g(k_j) \notin g^*(K_j))$.

Case I: $k_j \in K^*$.

As $ik_j \in g_1$ and $k_j \in K^*$, by definition of g^* we have $ik_j \in g^*$. Note that $\bar{m}(g', g_1) = m(g^*, g_1) > m(g^*, g) = \bar{m}(g', g)$. $ik_j \in g^*$ implies $g(i) \neq \varnothing$. $g + ik_j \in \mathrm{L}(g)$ implies $k_j \notin K$. Therefore $|g_1(K)| < |g(K)|$, so g_1 is unstable.

If $k_j \notin K^*$, then we have three cases.

Case IIa: $(g(i) \neq \varnothing, K \neq K^*)$ and $(\varnothing \neq g(k_j) \notin g^*(K_j))$.

Note that $\bar{m}(g', g_1) = m(g^*, g_1) \geq m(g^*, g) = \bar{m}(g', g)$. As $g(k_j)$ is single at g_1, g_1 is unstable.

Case IIb: $(g(i) \neq \varnothing, K \neq K^*)$ and $(\varnothing \neq g(k_j) \in g^*(K_j)$ or $g(k_j) = \varnothing)$.

By definition of g^*, it must be that $g(k_j) = g^*(k_j)$. Note that $m(g^*, g_1) \geq m(g^*, g) - 2$. $g_1(k_j) = i \neq g(k_j)$, so k_j must have strict preferences over g and g_1. Assumption 4 implies $k_j \notin K$, and $k_j \notin K^*$, $K \neq K^*$ by the assumptions of Case IIb, so i must have strict preferences over g_1, g and g^*.

[26] This is due to the main theorem of Roth [33]; for all $g, g^* \in \mathfrak{C}$, if $|g(K)| < q_K$, then $g(K) = g^*(K)$. That is, any college with unfilled places in some stable matching is matched to the same set of students in any stable matching. A corollary of this is that any college must be matched to the same number of students in any stable matching.

First, suppose that k_j prefers g, and therefore g^*, to g_1. If $g(k_j) \neq \emptyset$, let k_j and $g(k_j) = g^*(k_j)$ be matched, and if $g(k_j) = \emptyset$, let k_j leave i to become single. Denote the resulting network g_2. Note that $\bar{m}(g', g_2) \geq m(g^*, g_2) = m(g^*, g) = \bar{m}(g', g)$. As i is single at g_2, g_2 is unstable.

Next, suppose that k_j prefers g_1 to g and g^*. i must prefer g and g^* to g_1, otherwise (i, k_j) would be a blocking pair for g or g^*.

If i prefers g to g^*, then by Corollary B.1, $g^*(i)$ must prefer g^* to g (and to g_1, as $g_1(g^*(i)) = g(g^*(i))$). Let i and $g^*(i)$ be matched. Denote the resulting network g_3. Note that $\bar{m}(g', g_3) \geq m(g^*, g_3) \geq m(g^*, g) = \bar{m}(g', g)$. As $g_3(g(i)) = \emptyset$, $|g_3(K)| < |g(K)|$, so g_3 is unstable.

If i prefers g^* to g, then by Corollary B.1, $g(i)$ must prefer g to g^* and $g^*(g(i))$ must prefer g^* to g. Let $g^*(g(i))$ and $g(i)$ be matched. Denote the resulting network g_4. Note that $\bar{m}(g', g_4) \geq m(g^*, g_4) \geq m(g^*, g) = \bar{m}(g', g)$. If $g^*(g(i)) = g(k_j)$, then as $g(k_j) = g^*(k_j)$ (Case IIb), we have $g^*(g(i)) = g^*(k_j)$, which implies $g(i) = k_j$, contradicting $g_1 = g + ik_j \in \mathrm{L}(g)$. So, if $g(k_j) \neq \emptyset$, then $g(k_j)$ is single at g_4 and g_4 is unstable. If $g(k_j) = \emptyset$, then $g_4(i) = k_j \in K_j$ implies $g_4(K_j) \neq g(K_j)$, so by Roth [33], g_4 is unstable.

Case IIc: $(g(i) = \emptyset$ or $g(i) \neq \emptyset, K = K^*)$ and $(\emptyset \neq g(k_j) \notin g^*(K_j))$.

By definition of g^*, it must be that $g(i) = g^*(i)$. Note that $m(g^*, g_1) \geq m(g^*, g) - 2$. Assumption 4 implies $k_j \notin K$, so i must have strict preferences over g and g_1. $g_1 = g + ik_j \in \mathrm{L}(g)$ implies $g(k_j) \neq g_1(k_j)$, and $g^*(k_j) \neq i$, $g(k_j) \notin g^*(K_j)$ by the assumptions of Case IIc, so k_j must have strict preferences over g_1, g and g^*.

First, suppose that i prefers g, and therefore g^*, to g_1. If $g(i) \neq \emptyset$, let i and $g(i) = g^*(i)$ get matched. If $g(i) = \emptyset$, let i leave k_j to be single. Let g_5 denote the resulting network. Note that $\bar{m}(g', g_5) \geq m(g^*, g_5) \geq m(g^*, g) = \bar{m}(g', g)$. Since $g(k_j)$ is single, g_5 is unstable.

Next, suppose that i prefers g_1 to g and g^*. k_j must prefer g and g^* to g_1, otherwise (i, k_j) would be a blocking pair for g or g^*.

If k_j prefers g^* to g, then by Corollary B.1, $g(k_j)$ must prefer g to g^* and $g^*(g(k_j))$ must prefer g^* to g. Let $g^*(g(k_j))$ and $g(k_j)$ get matched. Let g_6 denote the resulting network. If $g(g^*(g(k_j))) = \emptyset$, then by definition of g^* and Roth [33] we have $g^*(g^*(g(k_j))) = g(k_j) = \emptyset$, contradicting our assumptions for Case IIc. If $g(g^*(g(k_j))) \neq \emptyset$, then $g(g^*(g(k_j)))$ is single at g_6, so g_6 is unstable.

If k_j prefers g to g^*, then by Corollary B.1, $g^*(k_j)$ must prefer g^* to g. Let k_j and $g^*(k_j)$ get matched. Let g_7 denote the resulting network. Note that $\bar{m}(g', g_7) \geq m(g^*, g_7) \geq m(g^*, g) = \bar{m}(g', g)$. Since $g(k_j)$ is single at g_7, g_7 is unstable.

Step 2: Given the g_T from Step 1, Lemma 5.5 implies the existence of $T_1 \in \mathbb{N}_+$, $g_{T_1} \in G$, such that $P_0^{T_1}(g_T, g_{T_1}) > 0$, $\bar{m}(g', g_{T_1}) > \bar{m}(g', g_T)$. If $g_{T_1} \in \mathfrak{C}$ then set $g'' = g_{T_1}$. Otherwise, apply Lemma 5.5 again to attain g_{T_2} such that $\bar{m}(g', g_{T_2}) > \bar{m}(g', g_{T_1})$. As $\bar{m}(g', .)$ is bounded above by $|N|$, we must eventually reach some $g_{T_m} \in \mathfrak{C}$ such that $\bar{m}(g', g_{T_m}) > \bar{m}(g', g)$. □

Proof of Theorem 5.8. Replacing $m(.,.)$ with $\bar{m}(.,.)$ and using Lemma B.3 instead of Lemma 3.3, the proof follows identical steps to the proof of Theorem 3.4. □

References

[1] A. Abdulkadiroğlu, Y.K. Che, Y. Yasuda, Resolving conflicting preferences in school choice: the "Boston mechanism" reconsidered, Amer. Econ. Rev. 101 (2011) 399–410.

[2] M. Agastya, Adaptive play in multiplayer bargaining situations, Rev. Econ. Stud. 64 (1997) 411–426.

[3] J. Bergin, B.L. Lipman, Evolution with state-dependent mutations, Econometrica 64 (1996) 943–956.

[4] P. Biró, M. Bomhoff, P.A. Golovach, W. Kern, D. Paulusma, Solutions for the stable roommates problem with payments, in: M. Golumbic, M. Stern, A. Levy, G. Morgenstern (Eds.), Graph-Theoretic Concepts in Computer Science, in: Lecture Notes in Computer Science, vol. 7551, Springer, Berlin/Heidelberg, 2012, pp. 69–80.

[5] P. Biró, G. Norman, Analysis of stochastic matching markets, Int. J. Game Theory 42 (2013) 1021–1040.

[6] L.E. Blume, The statistical mechanics of strategic interaction, Games Econ. Behav. 5 (1993) 387–424.

[7] J. Boudreau, An exploration into why some matchings are more likely than others, in: Proceedings of MATCH-UP 2012: the Second International Workshop on Matching Under Preferences, 2012, pp. 39–50.

[8] J.W. Boudreau, A note on the efficiency and fairness of decentralized matching, Oper. Res. Lett. 39 (2011) 231–233.

[9] B. Chen, S. Fujishige, Z. Yang, Decentralized market processes to stable job matchings with competitive salaries, KIER working papers, 2010.

[10] E. Diamantoudi, L. Xue, E. Miyagawa, Random paths to stability in the roommate problem, Games Econ. Behav. (2004) 18–28.

[11] E. Dokumaci, W.H. Sandholm, Large deviations and multinomial probit choice, J. Econ. Theory 146 (2011) 2151–2158.

[12] F. Echenique, L. Yariv, An experimental study of decentralized matching, Unpublished manuscript, California Institute of Technology, 2012.

[13] G. Ellison, Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution, Rev. Econ. Stud. 67 (2000) 17–45.

[14] A.M. Feldman, Recontracting stability, Econometrica 42 (1974) 35–44.

[15] M.L. Fredman, R.E. Tarjan, Fibonacci heaps and their uses in improved network optimization algorithms, J. ACM 34 (1987) 596–615.

[16] D. Gale, L.S. Shapley, College admissions and the stability of marriage, Am. Math. Mon. 69 (1962) 9–15.

[17] J.R. Green, The stability of Edgeworth's recontracting process, Econometrica 42 (1974) 21–34.

[18] D. Gusfield, R.W. Irving, The Stable Marriage Problem: Structure and Algorithms, The MIT Press, 1989.

[19] M.O. Jackson, A. Watts, The evolution of social and economic networks, J. Econ. Theory 106 (2002) 265–295.

[20] M. Kandori, G.J. Mailath, R. Rob, Learning, mutation, and long run equilibria in games, Econometrica 61 (1993) 29–56.

[21] B. Klaus, F. Klijn, M. Walzl, Stochastic stability for roommate markets, J. Econ. Theory 145 (2010) 2218–2240.

[22] B. Klaus, F. Payot, Paths to stability in the assignment problem, Cahiers de Recherches Economiques du Département d'Econométrie et d'Economie politique (DEEP), No. 13.14, Université de Lausanne, Faculté des HEC, DEEP, 2013.

[23] F. Kojima, M. Ünver, Random paths to pairwise stability in many-to-many matching problems: a study on market equilibration, Int. J. Game Theory 36 (2008) 473–488.

[24] M. Kwiatkowska, G. Norman, D. Parker, PRISM 4.0: Verification of probabilistic real-time systems, in: G. Gopalakrishnan, S. Qadeer (Eds.), Proc. 23rd International Conference on Computer Aided Verification, CAV'11, Springer, 2011, pp. 585–591.

[25] M. Maschler, B. Peleg, L.S. Shapley, Geometric properties of the kernel, nucleolus, and related solution concepts, Math. Oper. Res. 4 (1979) 303–338.

[26] D.P. Myatt, C. Wallace, A multinomial probit model of stochastic evolution, J. Econ. Theory 113 (2003) 286–301.

[27] R. Myerson, Refinements of the Nash equilibrium concept, Int. J. Game Theory 7 (1978) 73–80.

[28] H.H. Nax, B.S.R. Pradelski, Decentralized dynamics and equitable core selection in assignment games, Working paper, University of Oxford, 2013.

[29] J. Newton, Recontracting and stochastic stability in cooperative games, J. Econ. Theory 147 (2012) 364–381.

[30] G. Noldeke, L. Samuelson, An evolutionary analysis of backward and forward induction, Games Econ. Behav. 5 (1993) 425–454.

[31] J. Pais, A. Pinter, R.F. Veszteg, Decentralized matching markets: a laboratory experiment, Working papers, Department of Economics, University of Lisbon, 2012.

[32] A.E. Roth, The college admissions problem is not equivalent to marriage problem, J. Econ. Theory 36 (1985) 277–288.

[33] A.E. Roth, On the allocation of residents to rural hospitals: a general property of two-sided matching markets, Econometrica 54 (1986) 425–427.

[34] A.E. Roth, M.A.O. Sotomayor, Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis, first edition, Cambridge University Press, 1992.

[35] A.E. Roth, J.H. Vande Vate, Random paths to stability in two-sided matching, Econometrica 58 (1990) 1475–1480.

[36] W.H. Sandholm, Population Games and Evolutionary Dynamics. Economic Learning and Social Evolution, MIT Press, Cambridge, Mass, 2010.
[37] R. Sawa, Coalitional stochastic stability in games, networks and markets, Games Econ. Behav. 88 (2014) 90–111.
[38] R. Selten, Reexamination of the perfectness concept for equilibrium points in extensive games, Int. J. Game Theory 4 (1975) 25–55.
[39] L. Shapley, M. Shubik, The assignment game I: the core, Int. J. Game Theory 1 (1971) 111–130.
[40] H.P. Young, The evolution of conventions, Econometrica 61 (1993) 57–84.