A one-shot deviation principle for stability in matching problems

Jonathan Newton & Ryoji Sawa

November 2013
A one-shot deviation principle for stability in matching problems

Jonathan Newton†1 and Ryoji Sawa‡2

1School of Economics, University of Sydney
2Center for Cultural Research and Studies, University of Aizu

November 27, 2013

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.

Keywords: Learning; stochastic stability; matching; marriage; college admissions.

JEL Classification Numbers: C71, C72, C73, C78, D71.

†Come rain or come shine, I can be reached at jonathan.newton@sydney.edu.au.
‡Address: Tsuruga, Ikki-machi, Aizu-Wakamatsu City, Fukushima, 965-8580 Japan, telephone: 81-242-37-2500, e-mail: rsawa@u-aizu.ac.jp.


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 partnership. 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 (1962) as described in Jackson and Watts (2002). 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 an NTU version of the least-core as described in Maschler, Peleg and Shapley (1979). The results extend to one-sided matching markets (roommate problems) and to many-to-one matchings (college admissions problems). We apply our results to search for conditions on preferences which ensure stochastic stability of the preferred stable matching of one side of a two sided market.

In the related papers of Jackson and Watts (2002) and Klaus, Klijn and Walzl (2010), 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. A single mistake can be fatal to a partnership and can drive the dynamic process of partnership formation to a new equilibrium, from which in turn a single mistake can take the process elsewhere. The process can in this way move from any equilibrium to any other equilibrium. In the language of Noldeke and Samuelson (1993), 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 in the sense of Kandori, Mailath and Rob (1993); Young (1993). When mistakes are rare, in the long run the process will spend almost all of its time at stochastically stable matchings.

It should be remembered that one of the strengths of stochastic stability as implemented by Young (1993), as opposed to, for example, asymptotic stability or evolutionary stable strategies, is that stochastic stability can measure the resilience of equilibria to multiple shocks. This makes transitions between equilibria depend on cardinal payoffs, rather
than just ordinal payoffs. Another way of inducing such a dependence is to make error probabilities directly dependent on payoffs, such as in the logit choice model (Blume, 1993). We take the latter approach, although similar results can be derived by allowing the history of a partnership to change gradually in response to shocks.

Jackson and Watts (2002) give payoffs as cardinal, although their perturbed dynamic process is not sensitive enough for results to depend on the magnitude of payoff differences. Klaus et al. (2010) features the more standard set up of ordinal preferences. 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.

In the model of this paper, it is no longer the case that all stable matchings are stochastically stable, as they are in the models of Jackson and Watts (2002) and Klaus et al. (2010). Rather, the identity of stochastically stable states can depend on the payoffs of individual players. In dynamic processes with mistakes, any mistake by a player or pair of players has a cost, which can be interpreted as a measure of how bad the mistake is for those committing it. Mistakes with higher cost occur less often. We show that most of the time we can restrict attention to the least costly deviation from any stable matching, which for the logit choice rule equates to the deviation causing the lowest payoff loss to the deviating players. The stochastically stable matchings are shown to be contained within the set of matchings with the highest cost least cost deviations. This result extends to roommate problems with nonempty core and to college admissions problems with responsive preferences.

There is a growing literature which looks at equilibrium selection in matching problems (Biró and Norman, 2012; Boudreau, 2012; Echenique and Yariv, 2012; Pais, Pinter and Veszteg, 2012). Typically, these papers use simulation 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.

---

1It is not uncommon to assume cardinal preferences in the literature on matching problems. Abdulkadirou˘glu, Che and Yasuda (2011) 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 (1962).

2Boudreau (2011) 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.”
initial condition. In contrast, the current paper models the long run.

Algorithms for implementing a stable matching often select a matching which is optimal for one side of the market. This is true, for example, for hospital-intern matching in the United States (Roth, 1984; Roth and Peranson, 1999). A natural goal is then to find conditions under which such a matching is stochastically stable. Transitions between stable matchings are usually driven by the errors of players who are relatively insensitive to differences between prospective partners. If men are highly payoff insensitive relative to women, then, under the logic choice rule, easiest possible deviations always involve a man losing payoff.³ This in itself is insufficient to ensure that the woman-optimal matching is stochastically stable. The addition of a condition that men’s preferences be sufficiently concave in their ordinal payoff rankings ensures that at least some women achieve their woman-optimal partnerships. Still, this does not ensure that the woman-optimal matching is stochastically stable. A further normalization of payoffs, so that players use their worst payoff of any stable matching as a reference point, achieves this goal.

³All the statements in this paragraph can, of course, be restated with the sides of the matchings reversed.

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 (Diamantoudi, Xue and Miyagawa, 2004; Roth and Vande Vate, 1990). Another related literature is the literature on convergence to the core in cooperative games with transferable utility (Agastya, 1997; Feldman, 1974; Green, 1974; Newton, 2012). 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 (Biró, Bomhoff, Golovach, Kern and Paulusma, 2012; Chen, Fujishige and Yang, 2012; Nax and Pradelski, 2013; Shapley and Shubik, 1971). Of particular note is the work of Nax and Pradelski (2013), who adapt the results of the current paper to obtain selection within the core of the assignment game.

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 the problem of finding conditions under which the optimal matching for one side of the market is stochastically stable. 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 (2002). There is a set of players, $N$, which is divided into a set of men, $M = \{m_1, \ldots, m_k\}$, and a set of women, $W = \{w_1, \ldots, w_l\}$. 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 $ij \in g \iff ji \in g$. 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 $g$. $g(i) = \emptyset$ 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:

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

In a slight abuse of notation, we sometimes write $g(i) = j$ for $g(i) = \{j\}$. Let $\mu = \{(i, j) : (\exists g \in \mathcal{G} : ij \in g)\}$ 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 $u : G \rightarrow \mathbb{R}^{|N|}$. 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 $g(i) \neq \emptyset$, then $u_i(g) = u_i(\{ig(i)\})$, and if $g(i) = \emptyset$, then $u_i(g) = u_i(\emptyset)$. 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) $\exists i \in M, j \in W : u_i(g + ij) > u_i(g)$ and $u_j(g + ij) > u_j(g)$.

We denote the set of stable matchings by $\mathcal{C}$. 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.\textsuperscript{4} 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 \( F_{g^t}(\cdot) \) with full support on \( \mu \). 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 \( g \).

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. We define a perturbed blocking dynamic as a dynamic in which from time to time 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. The results of this paper apply for a large class of perturbed blocking dynamics, discussed in the following section. For expositional purposes, for the examples in the paper we use the well known and understood logit choice rule. Under the logit choice rule, the probability of a given mistake being made depends on the payoff loss incurred by the erring players. \( \eta > 0 \) is a parameter. The closer \( \eta \) is to zero, the lower the probability of mistakes.

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

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

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

\textsuperscript{4}Our unperturbed dynamic is essentially the same as those of Roth and Vande Vate (1990), Jackson and Watts (2002) and Klaus et al. (2010).
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

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

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.

The perturbed dynamic defines a Markov chain on \( G \). We denote the transition probabilities of the Markov chain with parameter \( \eta \) by \( P_\eta(\cdot, \cdot) \). That is, \( P_\eta(g, g') \) is the probability that \( g^{t+1} = g' \), given that \( g^t = g \). Taking the transition probabilities in the limit as \( \eta \to 0 \) gives \( P_0(\cdot, \cdot) \), which belongs to the class of unperturbed dynamics described in Section 2.2.

A pair of players must both agree to form a partnership, while a player can unilaterally abandon a partnership. The chain with \( \eta > 0 \) is aperiodic and irreducible, hence has a unique stationary distribution \( \pi_\eta \). Let \( \pi_0 = \lim_{\eta \to 0} \pi_\eta \). 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 (Young, 1993) and from any matching there exists a finite sequence of transitions under the unperturbed process that culminates in a stable matching being reached (Jackson and Watts, 2002; Roth and Vande Vate, 1990). Therefore, the only recurrent classes of the unperturbed process are the individual stable states. Therefore \( SS \subseteq \mathcal{C} \). 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:

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

We assume that, for all $g, g' \in G$, if $P_0(g, g') = 0$ and $P_\hat{\eta}(g, g') > 0$ for some $\hat{\eta} > 0$, equation (1) is well defined for some $c(g, g') \geq 0$. If $P_\eta(g, g') = 0$ for all $\eta \geq 0$, then let $c(g, g') = \infty$.

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:

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

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’ pay-offs for transition $g \to g'$ is the cost of $g \to g'$ (Sawa, 2013). We note that the results of every section aside from the examples in the paper and section 4 are not specific to the logit dynamic. Our results apply to every irreducible perturbed blocking dynamic which converges to a member of our class of unperturbed dynamics, for which $C(g, g')$ exists and is finite for all $g, g' \in G$, the limit $\pi_\eta \to \pi_0$ exists, and for which the cost of a mistake by a pair $(i, j)$ is independent of the current matching of every player other than $i$ and $j$.

A spanning tree rooted at $g^* \in \mathcal{C}$ is a directed graph over the set $\mathcal{C}$ such that every $g \in \mathcal{C}$ other than $g^*$ has exactly one exiting edge, and the graph has no cycles. 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 $g^* \in \mathcal{C}$ is stochastically stable if and only if there exists a minimum cost spanning tree rooted at $g^*$ (Young, 1993). Finding minimum cost spanning trees can be difficult.

The principal contribution of the current paper is to show that any stochastically stable matching, that is to say any root of a minimum cost spanning tree, 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:

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

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

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

c$_L(g)$ will be used to denote the cost of the least cost deviation from $g$.\(^9\)

$$c_L(g) := \min_{g' \neq g} c(g, g').$$

Note that for $g \notin \mathcal{C}$, there is a zero cost transition from $g$. We use the word deviation as we shall be interested in the application of these concepts to $g \in \mathcal{C}$. If the easiest transition at matching $g$ is for two players to form a partnership, then under the logit choice rule:

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

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

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

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

**Example 2.6** 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. For example, the top left cell tells us that $w_1$ gets a payoff of 30 from being matched with $m_1$. Assume $2 < a < 4$ and that payoffs from remaining unmatched are zero for every player. Let the perturbed dynamic be the logit choice rule.

\[
\begin{array}{ccc}
& w_1 & w_2 & w_3 \\
m_1 & 4,30 & 5,20 & a,30 \\
m_2 & 5,20 & 4,30 & a,20 \\
m_3 & 5,10 & 4,10 & a,10 \\
\end{array}
\]

\(^9\)This differs from the concept of the radius of a stable state $g \in \mathcal{C}$ (Ellison, 2000, citing a no longer extant working paper of Evans, 1993). The radius is defined as $R(g) = \min_{g' \in \mathcal{C} \setminus \{g\}} C(g, g')$ and requires a different stable state to be reached by the process. It turns out that in the problems considered in the current paper $c_L(g) = R(g)$ for all stable matchings outside of a specific set, but this does not follow from the definitions.
The stable matchings are \( g_W = \{m_1w_1, m_2w_2, m_3w_3\} \) and \( g_M = \{m_1w_2, m_2w_1, m_3w_3\} \). \( g_W \) is the woman optimal matching and \( g_M \) the man optimal matching. The least cost deviation from \( g_W \) has \( \{m_1w_3\} \) forming a link. Let \( g' \) denote the resulting matching. The cost of this deviation is:

\[
c_L(g_W) = c(g_W, g') = u_{m_1}(g_W) - u_{m_1}(g') = 4 - a
\]

This can be followed by \( \{m_2w_1\} \), \( \{m_1w_2\} \), \( \{m_3w_3\} \) forming links sequentially. These transitions have zero cost so the overall cost of a transition from \( g_W \) to \( g_M \) is \( C(g_W, g_M) = 4 - a \).

The least cost deviation from \( g_M \) has \( \{m_1, w_1\} \) forming a link. Let \( g'' \) denote the resulting matching. The cost of this deviation is:

\[
c_L(g_M) = c(g_M, g'') = u_{m_1}(g_M) - u_{m_1}(g'') = 5 - 4 = 1.
\]

This can be followed by \( \{m_2w_2\} \) forming a link. This transition has zero cost so the overall cost of a transition from \( g_M \) to \( g_W \) is \( C(g_M, g_W) = 1 \). The cost of the transition from \( g_W \) to \( g_M \) is greater than the opposite transition if \( a < 3 \) and lower if \( a > 3 \). Our main result in the next section will show that one can usually ignore all but the first step in each of these transitions: if a stable matching has a strictly higher \( c_L(\cdot) \) than all other stable matchings, then it is uniquely stochastically stable. Transitions subsequent to the initial deviation can be ignored. For this example, if \( a < 3 \), \( g_W \) is uniquely stochastically stable, and if \( a > 3 \), \( g_M \) is uniquely stochastically stable.

### 3 Stochastically stable matchings

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

\[
OS = \left\{ g \in G : c_L(g) = \max_{g' \in G} c_L(g') \right\}.
\]

As \( c_L(g) \) is strictly positive only for \( g \in \mathcal{C} \), it must be that \( OS \subseteq \mathcal{C} \). 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. In example 2.6 we see that \( c_L(g_W) = 4 - a \) and \( c_L(g_M) = 1 \). Therefore, if \( a < 3 \), then \( OS = \{g_W\} \), whereas if \( a > 3 \), then \( OS = \{g_M\} \).

Klaus et al. (2010) show that a single mistake suffices to move from any \( g \in \mathcal{C} \) to some other \( g' \in \mathcal{C} \). 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 sub-
sequently lead the process closer to $OS \subseteq \mathcal{C}$. This result is proved in Lemma 3.4, from which the main theorem is proven using a minimal cost spanning tree argument. First, we present a couple of lemmas which assist in the proof of Lemma 3.4.

The following lemma shows that the least cost deviation from a stable matching $g \notin OS$ cannot involve two single players forming a partnership.

**Lemma 3.1** Suppose that $g \in \mathcal{C}$ and $g \notin OS$. If $(i, j) \in N_L(g)$, then $g(i) \neq \emptyset$ and/or $g(j) \neq \emptyset$.

**Proof.** Suppose $g(i) = \emptyset$ and $g(j) = \emptyset$. Then, $i$ and $j$ are single in every stable matching (Theorem 2.22 of Roth and Sotomayor (1992)), including the matchings in $OS$. As $(i, j) \in N_L(g)$, $g + ij \in L(g)$. Then for $g^* \in OS$, $c_L(g^*) \leq c(g^*, g^* + ij) = c(g, g + ij) = c_L(g)$, therefore $g \in OS$, which contradicts our premise. ■

The next lemma demonstrates that if a pair is involved in a least cost deviation from a stable matching $g \notin OS$, then they do not both have the same current partner as in some matching within $OS$.

**Lemma 3.2** Suppose that $g \in \mathcal{C}$ and $g \notin OS$. If $(i, j) \in N_L(g)$, then for all $g^* \in OS$, $g(i) \neq g^*(i)$ and/or $g(j) \neq g^*(j)$.

**Proof.** Let $g^* \in OS$. Suppose $g(i) = g^*(i)$ and $g(j) = g^*(j)$. If $g(i) = j$, then $c_L(g^*) \leq c(g^*, g^* - ij) = c(g - ij) = c_L(g)$. If $g(i) \neq j$, then $c_L(g^*) \leq c(g^*, g^* + ij) = c(g, g + ij) = c_L(g)$. 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 3.3** $m(g, g')$ is the number of players who have the same partner in $g$ and $g'$.

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

**Lemma 3.4 (Getting Closer Lemma)** Let $g^* \in OS$. Suppose that $g \in \mathcal{C}$ and $g \notin OS$. Let $g_1 \in L(g)$. Then, $\exists g' \in \mathcal{C}, t \in \mathbb{N}_+, \text{ such that } m(g^*, g') > m(g^*, g) \text{ and } P_0^t(g_1, g') > 0$.

The proof is given in the appendix. The proof shows that from any $g \in \mathcal{C}, g \notin OS$, any least cost deviation leads to an unstable state $g_1 \notin \mathcal{C}$. Given some target matching $g^* \in OS$, starting from $g_1$, it is possible, under the unperturbed dynamic, to reach an unstable state, say $\tilde{g} \notin \mathcal{C}$, which is at least as close to $g^*$ as $g$ is to $g^*$. That is, $m(g^*, \tilde{g}) \geq m(g^*, g)$. As $\tilde{g}$ is unstable, the structure of the marriage problem ensures (see Lemma 5 of Klaus
et al., 2010 and Lemma 5.5 of the current paper) that, starting from $\tilde{g}$, it is possible, under the unperturbed dynamic, to reach a stable matching $g' \in \mathcal{C}$ which is strictly closer to $g^*$ than $\tilde{g}$ is to $g^*$. That is, $m(g^*,g') > m(g^*,\tilde{g})$. In combination, these inequalities give $m(g^*,g') > m(g^*,g)$. The least cost deviation from $g$ has sufficed to move the process to a stable matching which is closer to some $g^* \in OS$. Lemma 3.4 in hand, we can prove the main theorem.

**Theorem 3.5** $SS \subseteq OS$.

The formal proof is in the appendix. In brief, any stochastically stable matching must be the root of a minimum cost spanning tree. If a tree is rooted at some $g \in \mathcal{C}$, $g \notin OS$, then Lemma 3.4 can be used to build another tree rooted at some $g^* \in OS$. Starting at $g$, use Lemma 3.4 to add edges between stable matchings which get progressively closer to $g^*$. We obtain a sequence $(g = g_1, \ldots, g_L = g^*)$ with edges between $g_i$ and $g_{i+1}$ for $i = 1, \ldots, L - 1$. Each of these new edges has the cost of a lowest cost deviation, $C(g_i,g_{i+1}) = c_L(g_i)$. Deleting the edge exiting $g^*$, we are left with a tree rooted at $g^*$. As $g \notin OS$, $g^* \in OS$, the cost of the new edge exiting $g$ must be lower than the cost of the deleted edge which exited $g^*$. So the tree rooted at $g^*$ 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, $g \notin SS$.

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, as illustrated by the next example.

**Example 3.6** 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.\(^\text{10}\) The payoffs from being single are zero for all men and women. Let the perturbed dynamic be the logit choice rule.

\[
\begin{array}{ccc}
   & w_1 & w_2 & w_3 \\
m_1 & 10,1 & 5,5 & 1,10 \\
m_2 & 1,10 & 10,1 & 5,5 \\
m_3 & 5,5 & 1,10 & 10,1 \\
\end{array}
\]

\(^\text{10}\)This example is a version of Example 2.17 of Roth and Sotomayor (1992) in which we have removed a man and a woman.
There are three stable matchings as below.

\[ g_1 = \{m_1w_1, m_2w_2, m_3w_3\}, \quad g_2 = \{m_1w_2, m_2w_3, m_3w_1\}, \]

\[ g_3 = \{m_1w_3, m_2w_1, m_3w_2\}. \]

Observe that \( g_1 \) is man-optimal and \( g_3 \) is woman-optimal. It is easy to show that,

\[ c_L(g) = 1 \quad \text{for} \quad g \in \{g_1, g_3\}. \]

For example, one of the least cost deviations from \( g_1 \) is \( w_1 \) becoming single, which costs 1. Followed by \( \{m_1w_2\}, \{m_2w_3\}, \{m_3w_1\} \) forming links 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 \( g_2 \).

We can further characterize properties of matchings in \( OS \) for generic payoffs.

**Remark 3.7** For generic payoffs, the set of pairs of players involved in least cost deviations is a singleton, and is identical across matchings in \( OS \). That is, \( N_L(g) = N_L(g') \) for all \( g, g' \in OS \). This implies that \( g(i) = g'(i) \) and \( g(j) = g'(j) \) for \( (i, j) \in N_L(g) \) for all \( g, g' \in OS \).

Under the logit choice rule, \( OS \) corresponds to a non-transferable utility version of the least-core (Maschler et al., 1979).

**Remark 3.8** For \( A \subseteq N \), let \( G(A, g) \) be the set of matchings \( g' \neq g \) such that \( ij \in g' \) for all \( i, j \notin A \), \( ij \notin g \), and \( ij \notin g' \) for all \( i \notin A, ij \notin g \). Then the excess of \( A \) at \( g \) is defined as

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

and the least-core is

\[ \mathcal{LC} := \arg\min_{g \in \mathcal{C}} \max_{A \neq \emptyset} e(A, g). \]

In marriage problems, the maximum excess can always be found by analyzing \( A \) such that \( |A| \leq 2 \). For the logit choice rule, it then follows that \( OS = \mathcal{LC} \). Note that in contrast to the definition of excess in Maschler et al. (1979), 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.
Finally, we note that the proof of Theorem 3.5 extends to give a bound on convergence times.

**Remark 3.9** It follows immediately from the proof of Theorem 3.5 that the modified-coradius (see Ellison, 2000) of OS equals $\max_{g \in \text{OS}} c_L(g)$ and that therefore, starting from any matching, the expected hitting time of OS is $O(e^{\frac{1}{2} \max_{g \in \text{OS}} c_L(g)})$.

## 4 An application: Stochastic stability of woman-optimal matching

Many real-world applications of two-sided matching theory select a stable matching most preferred by one population. For example, the old algorithm used by the hospital-intern matching program in the U.S. selects the hospital-optimal matching, and the new algorithm selects the intern-optimal matching.\(^{11}\) However, the desired matching may not be stochastically stable, so even if implemented in the short run, it may be rarely observed in the long run. In this section, we consider sufficient conditions under which the unique stochastically stable matching is a matching optimal for one population. These conditions turn out to be strong. We choose to seek conditions under which the woman-optimal matching is stochastically stable. To simplify analysis, we assume that transitions between stable matchings are driven by errors by men. An alternative assumption with a similar effect would be to assume that men are considerably less sensitive to differences between prospective partners than are women.

**Assumption 1** Women do not make mistakes.

This assumption on its own does not predict that stable matchings which are better for any particular side will be chosen. $c_L(g)$ could still be low or high for any particular stable matching. However, it does lead to some interesting results. It is assumed throughout this section that the perturbed dynamic is the logit choice rule. To begin, the following lemma shows that if a least cost deviation involves a married man becoming single, then that man must be currently partnered with his worst partner of any stable matching.\(^{12}\)

**Lemma 4.1** Suppose $g \in \mathcal{C}$ and Assumption 1 holds. If $i \in M$ and $g - ig(i) \in L(g)$, then $g(i) = g_W(i)$.

---

\(^{11}\)There was a significant change in the program in 1998. We mean by the old algorithm the algorithm used until 1998 and by the new one the algorithm used since then. See Roth (1984) for details of the old algorithm and Roth and Peranson (1999) for the new one.

\(^{12}\)It follows that only players who are already relatively unhappy will make the mistake of returning to single life.
Proof. Suppose $g(i) \neq g_W(i)$. Then, as $g_W(i)$ prefers $i$ to $g(g_W(i))$, $i$ prefers $g_W(i)$ to remaining single, and Assumption 1 holds, a deviation from $g$ to $g + ig_W(i)$ must have lower cost than a deviation to $g - ig(i)$, contradicting its being a least cost deviation. ■

The following lemma shows that if a least cost deviation has any man leaving his current partner for a currently single woman, he must currently be in his least preferred stable matching.\(^\text{13}\)

**Lemma 4.2** Suppose that $g \in \mathcal{C}$ and Assumption 1 holds. Suppose $i \in M$, $g(i) \neq \varnothing$, $g(j) = \varnothing$. If $(i,j) \in N_L(g)$, then $g(i) = g_W(i)$.

Proof. Suppose $i \in M$ and $g(i) \neq g_W(i)$. Then, $g_W(i)$ prefers $i$ to $g(g_W(i))$. As women do not make mistakes (assumption 1) it must be the case that $j$ prefers $i$ to remaining single. $i$ must prefer $g_W(i)$ to $j$, otherwise $(i,j)$ would be a blocking pair for $g_W$. But then a deviation from $g$ to $g + ig_W(i)$ must have lower cost than a deviation to $g + ij$, contradicting its being a least cost deviation. ■

The next lemma shows that if a man and a woman participate in a least cost deviation, and the man is not already with his least preferred partner of any stable matching, then the woman is not already with her most preferred partner of any stable matching.\(^\text{14}\)

**Lemma 4.3** Suppose that $g \in \mathcal{C}$ and Assumption 1 holds. Suppose $i \in M$, $j \in W$. Let $(i,j) \in N_L(g)$. If $g(i) \neq g_W(i)$, then $g(j) \neq g_W(j)$.

Proof. If $j = g_W(i)$, then $g(j) \neq g_W(j)$ and we are done. $j$ prefers $i$ to $g(j)$ by assumption 1. If $j \neq g_W(i)$, it must be the case that $i$ prefers $j$ to $g_W(i)$, as $(i,j) \in N_L(g)$. So, if $g(j) = g_W(j)$, then $(i,j)$ would be a blocking pair for $g_W$, contradicting its being a stable matching. ■

### 4.1 A specific utility function

In this section we assume that players are much more sensitive to differences between partners close to the bottom of their ordinal ranking of prospective partners. They are more sensitive to ugliness than to beauty.\(^\text{15}\) Let $x_i(\_): G \rightarrow \mathbb{Z}$ denote an ordinal ranking (low numbers being worse) of $g \in G$ in player $i$’s preferences. That is, for $g \in G$: if $g'$ is such that $u_i(g') > u_i(g)$ and there does not exist $g''$ such that $u_i(g') > u_i(g'') > u_i(g)$, then

\(^{13}\)It follows that if you are concerned about your partner making a mistake and quitting you, unless he or she is already relatively unhappy, you should not be concerned about single rivals.

\(^{14}\)Similarly to the previous footnote, neither should you be concerned about rivals partnered with their best partner of any stable matching.

\(^{15}\)To quote the Marquis de Sade: “Beauty belongs to the sphere of the simple, the ordinary, whilst ugliness is something extraordinary, and there is no question but that every ardent imagination prefers in lubricity, the extraordinary to the commonplace.”
$x_i(g') = x_i(g) + 1$. The utility function will convert ordinal preferences to cardinal utility values in a particular way.\footnote{An assumption that $\beta_M \ll \beta_F$ would ensure that men are considerably less sensitive to differences between prospective partners than are women. As noted above, this could replace assumption 1.}

**Assumption 2**

$$\forall i \in M, \quad u_i(g) = v(x_i(g)) = \beta_M \left(1 - e^{-x_i(g)} \right),$$

$$\forall i \in F, \quad u_i(g) = v(x_i(g)) = \beta_F \left(1 - e^{-x_i(g)} \right).$$

Note that any decrease in utility due to a rematching is always greater than the possible gain from any rematching. This is due to the functional form of the utility: it is very concave. Due to this, any given man will now be much more likely to make mistakes when matched to a better partner than when matched to a worse one.

Denote the set of men who can participate in a least cost deviation from $g$ by $I^*(g)$.

$$I^*(g) := \{i \in M : \exists (i, j) \in N_L(g) \}$$

Let $i^*(g) \in I^*(g)$ denote a representative man. Let $\tilde{G}$ be the set of matchings in which at least one of the men in $I^*(g)$ has the same partner as in $g_W$.

$$\tilde{G} := \{g \in C : g(i^*(g)) = g_W(i^*(g)) \text{ for some } i^*(g) \in I^*(g) \}.$$ 

Note that $g_W \in \tilde{G}$ by definition.

**Theorem 4.4** If Assumptions 1 and 2 are satisfied, then $OS \subseteq \tilde{G}$.

**Proof.** We show that, for all $i \in M$ and $g \in C$, some deviation involving $i$ from $g$ is weakly less costly than any deviation involving $i$ from $g_W$. Then, we show that $g \not\in OS$ if $g \not\in \tilde{G}$.

For any $i \in M$, $g \in C$, such that $g(i) \neq g_W(i)$, the least costly of all deviations involving $i$ from $g$, that is a deviation to $g - ig(i)$ or to $g + ij$ for some $j \in W$, has cost bounded above by the cost of a deviation to $g + ig_W(i)$:

$$c(g, g + ig_W(i)) = u_i(g) - u_i(g_W) < \beta_M - \beta_M \left(1 - e^{-x_i(g_W)} \right) = \beta_M e^{-x_i(g_W)}.$$ 

Any deviation involving $i \in M$ from $g_W$, that is a deviation to $g_W - ig_W(i)$ or to $g_W + ij$ for some $j \in W$, has cost bounded below by:

$$u_i(g_W) - \max_{g \in G} \{u_i(g) : u_i(g) < u_i(g_W)\} = v(x_i(g_W)) - v(x_i(g_W) - 1).$$
\[
\begin{align*}
&= \beta_M \left( 1 - e^{-x_i(g_w)} \right) - \beta_M \left( 1 - e^{-x_i(g_w) + 1} \right) = \beta_M e^{-x_i(g_w)} + 1 - \beta_M e^{-x_i(g_w)} \\
&= \beta_M e^{-x_i(g_w)} (e - 1).
\end{align*}
\]

So for \( i \in M, g(i) \neq g_W(i) \), there exists a deviation from \( g \) involving \( i \) which is strictly less costly than any deviation from \( g_W \) which involves \( i \).

Now consider deviations which involve \( i \in M \) for whom \( g(i) = g_W(i) \). As every \( j \in W \) weakly prefers \( g_W \) to \( g \), we have for all \( i \in M \),

\[
\{ j \in W : u_j(\{ij\}) > u_j(g) \} \supseteq \{ j \in W : u_j(\{ij\}) > u_j(g_W) \}.
\]

This implies that every deviation from \( g_W \) which involves \( i \in M \), \( g(i) = g_W(i) \), is also possible from \( g \). So for \( i \in M \), \( g(i) = g_W(i) \), there exists a deviation from \( g \) involving \( i \) which is weakly less costly than any deviation from \( g_W \) which involves \( i \).

The arguments above show that \( c_L(g) \leq c_L(g_W) \). Now, assume \( g \notin \tilde{G} \). Then \( g(i^*(g)) \neq g_W(i^*(g)) \) for all \( i^*(g) \in I^*(g) \). There therefore exist deviations from \( g \) involving \( i^*(g) \in I^*(g) \) which are strictly less costly than any deviation from \( g_W \). As the cost of deviations by these players is binding for the determination of \( c_L(g) \), it must be that \( c_L(g) < c_L(g_W) \). Therefore \( g \notin OS \). \( \blacksquare \)

So we have that if players on one side of the market are insensitive to payoff differences compared to those on the other side (captured by assuming that the other side do not make errors), and players are more sensitive to differences between prospective partners lower down their preference ordering, then insensitive players will usually do badly in stochastically stable matchings.\(^{17}\) However, note that if \( m_1 \) ranks \( w_1, w_2 \) at the top of his preference ordering, \( w_1 \) and \( w_2 \) rank \( m_1 \) at the top of their preference orderings, and \( m_1 \) likes \( w_1 \) as much as any other man likes any other woman, then for any \( g \in C \) we have that \( m_1 \in I^*(g), g(m_1) = w_1 \), and \( g + m_1 w_2 \in L(g) \). That is, \( \tilde{G} = OS = C \). True love and close competition make it easy to comment on how the lovers match in any stochastically stable matching, but make it difficult to comment on the behavior of everybody else. This is a bigger problem than it might seem, as although \( g_W \) is always in \( \tilde{G} \), \( g_W \) is not necessarily stochastically stable. That is, it is possible that \( \pi_0(\tilde{G}) > 0 \) and \( \pi_0(g_W) = 0 \).

**Example 4.5** Suppose that \( M = \{m_1, m_2, m_3\}, W = \{w_1, w_2, w_3, w_4\} \), and that the matrix giving their ordinal preference ranking from a given match is given in the left table of figure 1. If player \( i \) is unmatched in \( g \), then \( x_i(g) = 0 \). The conversion of ordinal to cardinal payoffs is as in the preceding section. For example, the top left cell tells us that \( w_1 \) gets a payoff of \( \beta_W (1 - e^{-3}) \)

\(^{17}\)In fact, the existence of even a single player who is very insensitive compared to every other player is enough to give this result.
from being matched with \( m_1 \). We assume Assumption 1 still holds.

\[ g_W = \{m_1w_1, w_2, m_3w_3, m_2w_4\} \text{ and } g_M = \{m_1w_1, w_2, m_3w_4, m_2w_3\} \] are the stable matchings. Note that \( m_1 \in I^*(g_W) \cap I^*(g_M) \), \( g_W(m_1) = g_M(m_1) \), \( g_W + m_1w_2 \in L(g_W) \), and \( g_M + m_1w_2 \in L(g_M) \). Therefore \( \{g_W, g_M\} = \tilde{G} \). A least cost transition from \( g_W \) to \( g_M \) starts with \( \{m_1w_2\} \) forming a link. The cost of this deviation is:

\[
c(g_W, g_W + m_1w_2) = u_{m_1}(g_W) - u_{m_1}(g_W + m_1w_2) = v(2) - v(1) = e^{-1} - e^{-2}
\]

This can be followed by \( \{m_3w_1\}, \{m_3w_4\}, \{m_1w_1\}, \{m_2w_3\} \) forming links sequentially. These transitions have zero cost so the overall cost \( C(g_W, g_M) = e^{-1} - e^{-2} \).

No least cost transition from \( g_M \) to \( g_W \) can commence with such a low cost deviation. \( m_2 \) and \( m_3 \) cannot make such a low cost deviation by matching with \( w_1 \) in error, as \( w_1 \) prefers to remain matched to \( m_1 \). Starting a path by \( \{m_1w_2\} \) forming a link leads to \( g_M + m_1w_2 \), from which no zero cost path can lead to \( g_W \). \( m_2 \) and \( m_3 \) are unwilling to leave their existing partners, and eventually \( \{m_1, w_1\} \) can rematch at zero cost to return the process to \( g_M \). Therefore the transition from \( g_M \) to \( g_W \) must have a higher cost than the reverse, and \( g_M \) is uniquely stochastically stable.

### 4.2 A sufficient condition on preferences

Sufficient conditions for the stochastic stability of \( g_W \) can be achieved with an additional condition: a normalization. Take the worst possible stable outcome for the men to be a reference point in the sense that they all receive the same utility in such a matching: if \( u_i(g_W) \) does not vary across \( i \in M \), then we can give a sharper characterization. In fact, under this assumption, the unique stochastically stable matching is the woman-optimal matching.

**Assumption 3** \( x_i(g_W) = 0 \) for all \( i \in M \).

**Theorem 4.6** If Assumptions 1, 2 and 3 are satisfied, then \( \text{OS} = \{g_W\} \).
Proof. Similarly to the proof of Theorem 4.4,
\[ c_L(g_W) \geq \min_{i \in M} \beta_M e^{-x_i(g_W)}(e - 1) = \beta_M (e - 1) > \beta_M \]

Suppose \( g \in C, g \neq g_W \). There exists \( i \) such that \( g(i) \neq g_W(i) \). Note that \( g_W(i) \) prefers \( g_W \) to \( g \), so \( c_L(g) \) must be bounded above by \( c(g, g + ig_W(i)) \), i.e.
\[ c_L(g) \leq c(g, g + ig_W(i)) = u_i(g) - u_i(g_W(i)) < \beta_M e^{-x_i(g_W)} = \beta_M. \]
Therefore \( c_L(g) < c_L(g_W) \) and \( g \not\in OS \). ■

The normalized payoffs from the example in section 4.5 are given in the right table of figure 1. OS now contains \( g_W \) only.

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 agent of one population, the colleges, may be matched with more than one agent of the other population, the students. Each student is matched with at most one college. There are two sets, \( K = \{K_1, \ldots, K_l\} \) 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 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:
\[ G = \{g \in G : (\forall ij \in g, i \in S \iff j \in K), (\forall i \in S, |g(i)| \leq 1), (\forall K_j \in K, |g(K_j)| \leq q_{K_j}\}. \]

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: \( g(K) \neq g'(K) \Rightarrow u_K(g) \neq u_K(g') \).

Definition 5.1 A matching \( g \) is in the core of a matching problem, denoted \( g \in C' \) if \( \exists A \subseteq N, g' \in G \) such that:
(i) \( i \not\in A, j \not\in A, ij \in g \Rightarrow ij \in g' \)
(ii) \( ij \not\in g, ij \in g' \Rightarrow i \in A, j \in A \)
(iii) \( i \in A \Rightarrow u_i(g') > u_i(g) \).
We restrict our attention to responsive preferences (Roth, 1985). 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 the college prefers $s_i$ to $s_j$, and $T$ 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 $u_K(\{Ks_i\}) > u_K(\{Ks_j\}) \iff u_K(g) > u_K(g')$.

(II) If $g(K) = g'(K) \cup \{s_i\}, s_i \notin g'(K)$, then $u_K(\{Ks_i\}) > u_K(\emptyset) \iff u_K(g) > u_K(g')$.

Following Chapter 5 of Roth and Sotomayor (1992), 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. With a slight abuse of notation, we let $K$ denote the set of positions in college $K$, i.e. $K = \{k_1, \ldots, k_{q_K}\}$, $g(K) = \bigcup_{1 \leq i \leq q_K} g(k_i)$.

We assume that both the unperturbed and the perturbed dynamics forbid positions of the same college from competing for students. In the original problem, college $K$ is indifferent between a student filling position $k_i \in K$ or $k_j \in K$. Hence it is unrealistic that two positions within the same college compete with one another for a student. We prevent such competition by imposing the assumption below.\(^{18}\)

**Assumption 4** Let $\nu(g) := \{(i, k_j) : k_j \in K \in \mathbf{K}, i \neq g(k_j), i \in g(K)\}$. At the beginning of period $t + 1$, the updating pair of players will be chosen according to a distribution with full support on $\mu \setminus \nu(g')$.

The next lemma shows that the set of absorbing states, $\mathcal{C}$, of the dynamic in section 2.2 amended to satisfy Assumption 4 corresponds to the core, $\mathcal{C}'$, in the college admission problem.

**Lemma 5.3** For $g' \in G$, let $g \in G$ satisfy: for all $K = \{k_1, \ldots, k_{q_K}\} \in \mathbf{K}$,

(a) $i \in g'(K) \iff \exists k_j \in K : g(k_j) = i$.

\(^{18}\)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.
(b) ∀i ∈ N, |g(i)| ≤ 1.

Then g' ∈ C' ⇔ g ∈ C.

Note that the logit dynamic under Assumption 4 is still irreducible. For any network g, a pair (g(k), k) will cut their link with positive probability in the perturbed dynamic satisfying Assumption 4. For all g, g' ∈ G, there is a positive probability that the perturbed dynamic starting with g will become the empty network within |N| periods, and then will become g' within another |N| periods.

**Definition 5.4** Define the set of matchings equivalent to g as:

\[ Eq(g) = \{g' ∈ G : g'(K) = g(K) \forall K ∈ K\} \]

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 admission problem.

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 labelling of the positions of any given college.

**Assumption 5** If ̃g ∈ Eq(g); k_1, k_2 ∈ K_i : g(k_1) = ̃g(k_2); s ∈ S : g(s), ̃g(s) ∈ K_j ∈ K or g(s) = ̃g(s) = ∅; then:

(i) c(g, g + k_1s) = c(̃g, ̃g + k_2s),

(ii) If g(k_1) ≠ ∅, then c(g, g − k_1g(k_1)) = c(̃g, ̃g − k_2̃g(k_2)), and

(iii) If g(s) ≠ ∅, then c(g, g − sg(s)) = c(̃g, ̃g − ̃s̃g(s)).

Note that the logit choice rule satisfies Assumption 5.

Take any unstable matching g ∉ C, and a target stable matching g' ∈ C. The following lemma, which is important to the results of the paper, 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. (2010) to many-to-one matching problems. First, define a similarity function for the many-to-one matching problem:

\[ \bar{m}(g, g') := \max_{\hat{g} ∈ Eq(g')} m(g, \hat{g}) \] (5)

Note that \( \bar{m}(g, g') \geq m(g, g') \). Also note that \( m(.,.) \equiv \bar{m}(.,.) \) for one-to-one matching problems.
Lemma 5.5 Assume Assumption 4 holds. Let $g \not\in \mathcal{C}$, $g' \in \mathcal{C}$. Then, $\exists T \in \mathbb{N}^+, g_T \in G$, such that $P_T^0(g, g_T) > 0$ and $\bar{m}(g_T, g') > \bar{m}(g, g')$.

The proof is left to the appendix, 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 $g \not\in \mathcal{C}$, $g' \in \mathcal{C}$. Let $g^* \in \text{argmax} \ m(g, \hat{g})$. 

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 $g^{**} \in \text{Eq}(g^*) = \text{Eq}(g')$ and $m(g, g^{**}) \geq m(g, g^*) + 2$, 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 (1990) except that we have not assumed students to have strict preferences over the positions within colleges.\footnote{See Chapter 5 of Roth and Sotomayor (1992) for a way to construct strict preferences in such problems.}

Corollary 5.7 (Random paths to stability) Suppose a college admission problem, its related marriage problem, and an unperturbed dynamic satisfying Assumption 4. For any $g \not\in \mathcal{C}$, there exists $T \in \mathbb{N}^+, g^* \in \mathcal{C}$, such that $P_T^0(g, g^*) > 0$.

Define $c_L(g)$ and $OS$ as in the one-to-one matching problem. Using Lemma 5.5, a many-to-one version of Lemma 3.4 can be proved. Then, we have the following theorem. See Appendix for proofs.

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

The implications of Assumption 4 for the costs of deviations under the logit choice rule are as follows. Let

$$ E(g) = \{ik_j : ik_j \not\in g, ik_j \in g' \text{ for some } g' \in \text{Eq}(g)\}. $$
In words, \( ik_j \in E(g) \) means that \( i \) and \( k_j \in K \) are not matched in \( g \), but \( i \) is matched with some \( k_i \in K, k_i \neq k_j \). Assumption 4 implies that pairs in \( E(g) \) may not deviate if the process is in \( g \). For the logit dynamic, expressions for \( c_L(g) \) will be as in expressions (3) and (4), but with the minimum being taken over \( ij \notin g \cup E(g) \) in expression (3). The next example shows an application of Theorem 5.8 with a note emphasizing the role of Assumption 4.

**Example 5.9** Let \( S = \{s_1, s_2, s_3\} \), \( K = \{K, K'\} \), \( K = \{k_1, k_2\} \) 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.

<table>
<thead>
<tr>
<th></th>
<th>( k_1 )</th>
<th>( k_2 )</th>
<th>( k_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>10,10</td>
<td>10,10</td>
<td>5,10</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>10,8</td>
<td>10,8</td>
<td>6,8</td>
</tr>
<tr>
<td>( s_3 )</td>
<td>5,9</td>
<td>5,9</td>
<td>10,5</td>
</tr>
</tbody>
</table>

Observe that under Assumption 4, the set of stable matchings is \( \mathcal{C} = \{g_1, g_2, g_3, g_4\} \) where

\[
\begin{align*}
  g_1 &= \{(s_1, k_1), (s_2, k_2), (s_3, k_3)\}, \\
  g_2 &= \{(s_1, k_2), (s_2, k_1), (s_3, k_3)\}, \\
  g_3 &= \{(s_1, k_1), (s_2, k_3), (s_3, k_2)\}, \\
  g_4 &= \{(s_1, k_2), (s_2, k_3), (s_3, k_1)\}.
\end{align*}
\]

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

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 \( \mathcal{C} \).

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 \( L(g_1) = \{g_1 + s_2k_3\} \) with cost \( c_L(g_1) = 4 \). Also, \( c_L(g_2) = 4, c_L(g_3) = 1, c_L(g_4) = 1 \). \( \text{OS} = \{g_1, g_2\} \). 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.

### 6 Roommate problems

In the one-sided matching problem, or roommate problem, the set of admissible matchings is not restricted to be bipartite. Anyone can partner with anyone. The set
of networks of interest is broadened to:

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

Gusfield and Irving (1989) 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 \( g, g' \in \mathcal{C}, i \) 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. (2004) show that if \( \mathcal{C} \) is nonempty, then there exists a sequence of mutually beneficial blockings ending in \( \mathcal{C} \). In the context of this paper, this means that nonempty \( \mathcal{C} \) implies that all recurrent classes of the unperturbed Markov process lie in \( \mathcal{C} \). There are no absorbing cycles. Assuming nonemptiness of \( \mathcal{C} \), lemmas 3.1, 3.2, 3.4 still hold. It follows that:

**Theorem 6.1** If \( \mathcal{C} \neq \emptyset \), 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 admission 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 complementarities, are left for future work.
A Appendix

In this section, we prove Theorem 5.8. Theorem 3.5 is implied by Theorem 5.8, as Assumptions 4 and 5 do not have any effect in the one-to-one setting. Similarly, Lemma 3.4 is implied by its many-to-one equivalent, Lemma A.4. The proof of Lemma 5.5, one of our key lemmas, is given as below.

**Proof of Lemma 5.5.** Let

\[ g^* \in \arg\max_{\hat{g} \in \mathcal{E}_l(g')} m(g, \hat{g}). \]

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

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

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 \( \tilde{m}(g_T, g') \geq m(g_T, g^*) > m(g_1, g^*) = m(g, g^*) = \tilde{m}(g, g') \) 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 \( \tilde{m}(\cdot, g^*) \).

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 \( i \in S \): \( g_1(i) \in K, g^*(i) \in K^*, K \neq K^* \), either \( u_i(g_1) > u_i(g^*) \) or \( u_i(g^*) > u_i(g_1) \). We assume that \( u_i(g_1) > u_i(g^*) \). The arguments when the converse holds are identical.

Let \( f : N \rightarrow N \) be such that \( f(j) = g_1(j) \) if \( u_i(g_1) > u_i(g^*) \) and \( f(j) = g^*(j) \) otherwise.\( ^{20} \)

Suppose a sequence \( \{i, f(i), f^2(i), f^3(i), \ldots \} \) where \( f^2(i) = f(f(i)) \) and \( f^k(\cdot) \) for \( k \geq 3 \) is defined similarly. Since \( N \) is finite, the sequence must repeat and create a cycle. Denote 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

\( ^{20} \) Note that strict preferences and the definition of \( g^* \) imply that \( u_j(g) \neq u_j(g^*) \) if \( g(j) \neq g^*(j) \).
and $g^*$, i.e. $g^*(n_j) = n_{j+1}$ if $j$ is odd, and $g_1(n_j) = n_{j+1}$ otherwise. \footnote{21} 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 $\hat{h}(i_k,j_k) \circ (g_1,g_2 = g_1 + ik_j) = 0$, $u_i(g_1) \geq u_i(g^*)$ and $u_{k_i}(g_1) \geq u_{k_i}(g^*)$. If there did exist such a $(i_k,j_k)$, then $u_i(g_2) > u_i(g_1) \geq u_i(g^*)$ and $u_{k_i}(g_2) > u_{k_i}(g_1) \geq u_{k_i}(g^*)$, so $(i_k,j_k)$ would be a blocking pair for $g^* \in \mathcal{C}$. So, $u_i(g^*) > u_i(g_1)$ and/or $u_{k_i}(g^*) > u_{k_i}(g_1)$. Without loss of generality, let $n_2$ be a member of a blocking pair for $g_1$ such that $u_{n_2}(g^*) > u_{n_2}(g_1)$. \footnote{22} Note that $g_2(g_1(n_2)) = g_2(n_1) = \varnothing$, and that $u_{n_m}(g^*) > u_{n_m}(g_1)$. $(n_1,n_m)$ is a blocking pair for $g_2$. Let $g_3 = g_2 + n_m$. $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 $\bar{m}(g_3,g') > m(g_3,g^*) = m(g,g^*)$ 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 $g_3(n_{m-1}) = \varnothing$, $u_{n_{m-2}}(g^*) > u_{n_{m-2}}(g_1)$. Let $g_4 = g_3 + n_{m-1}$. 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 $g_5 = g_3 + n_{m-3}$. 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 $n_{m-1}$ and $g_2(n_2)$ are positions in the same college. Therefore $\bar{m}(g_2,g') = \bar{m}(g_1,g')$. 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 $g,g^* \in \mathcal{C}$ 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 A.1 Let $g,g' \in \mathcal{C}$. Let $g^* \in \arg\max_{g \in \mathcal{C}} m(g,g)$. 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, 3.2 and 3.4. The next lemma is analogous to Lemma 3.1.

Lemma A.2 Suppose that $g \in \mathcal{C}$ and $g \notin OS$. If $i \in S, k_j \in K \in K, (i,k_j) \in N_L(g)$, then $g(i) \neq \varnothing$ and/or $g(k_j) \neq \varnothing$. 

\footnote{21}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 $g_1$.

\footnote{22}Such a member must be in a cycle since players who are not in cycles are indifferent between $g_1$ and $g^*$.}
Proof. Suppose \( g(i) = \emptyset \) and \( g(k_j) = \emptyset \). Then, in any \( g' \in \mathcal{C} \), \( g'(i) = \emptyset \) and there exists \( k_l \in K \) such that \( g'(k_l) = \emptyset \). As \((i, k_j) \in \mathcal{N}_L(g), g + ik_j \in L(g)\). Let \( g^* \in \mathcal{OS} \subseteq \mathcal{C} \), \( k_l \in K \) such that \( g^*(k_l) = \emptyset \). Then \( c_L(g^*) \leq c(g^*, g^* + ik_l) = c(g, g + ik_j) = c_L(g) \). Therefore \( g \in \mathcal{OS} \), which contradicts our premise. ■

The next lemma is analogous to Lemma 3.2.

Lemma A.3 Suppose that \( g \in \mathcal{C} \), \( g \not\in \mathcal{OS} \), and Assumption 4 holds. Suppose that \((i, k_j) \in \mathcal{N}_L(g)\). Let \( K, K' \in K \) be such that \( g(i) \in K \) and \( k_j \in K' \). Then, for all \( g^* \in \mathcal{OS} \), either \( g^*(i) \not\in K \) and/or \( g(k_j) \not\in g^*(K') \).

Proof. Let \( g^* \in \mathcal{OS} \). Suppose \( g^*(i) \in K \) and \( g(k_j) = g^*(k_l) \) for some \( k_l \in K' \). If \( K = K' \), then by Assumption 4, \( i = g(k_j) \), so \( i = g^*(k_l) \) and \( c_L(g^*) \leq c(g^*, g^* - ik_l) = c(g, g - ik_j) = c_L(g) \). If \( K \neq K' \), then \( c_L(g^*) \leq c(g^*, g^* + ik_l) = c(g, g + ik_j) = c_L(g) \). Therefore \( g \in \mathcal{OS} \), which contradicts our premise. ■

Lemma A.4 (Getting Closer Lemma II) Suppose the dynamic satisfies Assumptions 4, 5. Let \( g' \in \mathcal{OS} \). Suppose that \( g \in \mathcal{C} \) and \( g \not\in \mathcal{OS} \). Let \( g_1 \in L(g) \). Then, \( \exists g'' \in \mathcal{C}, t \in \mathbb{N}_+, \text{ such that } \tilde{m}(g', g'') > \tilde{m}(g', g) \) and \( P_0^T(g_1, g'') > 0 \).

Proof. Let \( g^* \) satisfy:

\[
g^* \in \arg\max_{g^* \in \mathcal{OS}} m(g, g^*)
\]

and

\[
g^* \in \arg\max_{g^* \in \mathcal{OS}} m(g_1, g^*).
\]

It is possible to choose such a \( g^* \) as any student matched to the same college in \( g \) and \( g_1 \) is matched to the same position of that college.

Suppose that \( g - ig(i) \in L(g) \). Suppose that \( i \in S \). Let \( g(i) \in K \) and \( g^*(i) \in K^* \). Under Assumption 5, \( g' \in \mathcal{OS} \) implies \( g^* \in \mathcal{OS} \). This, and \( g \not\in \mathcal{OS} \) imply \( g(i) \in K \neq K^* \subseteq g^*(i) \), so \( m(g^*, g_1) = m(g^*, g) \), and as \( \tilde{m}(g', g_1) = m(g^*, g_1) \) and \( \tilde{m}(g', g) = m(g^*, g) \), we have \( \tilde{m}(g', g_1) = \tilde{m}(g', g) \). Since \( g_1 \) is unstable (\( i \) is single), Lemma 5.5 guarantees there exists \( T \in \mathbb{N}_+, g_T \in \mathcal{G} \), such that \( P_0^T(g_1, g_T) > 0 \) and \( \tilde{m}(g', g_T) > \tilde{m}(g', g_1) = \tilde{m}(g', g) \).

Next, suppose that \( g + ik_j \in L(g) \). If \( g(i) \neq \emptyset \), let \( K \) be such that \( g(i) \in K \) and \( K^* \) be such that \( g^*(i) \in K^* \). Let \( k_j \in K_j \). Lemma A.3 implies that \((g(i) \neq \emptyset \) and \( K \neq K^* \)) and/or \((\emptyset \neq g(k_i) \not\in g^*(K_j))\). Furthermore, Lemma A.2 implies that \( g(i) \neq \emptyset \) and/or \( g(k_j) \neq \emptyset \).
Case I: \((g(i) \neq \emptyset, K \neq K^* \text{ and } \emptyset \neq g(k_j) \notin g^*(K_j))\) and/or \((ik_j \notin g^*)\).

Note that \(\tilde{m}(g',g_1) = m(g^*,g_1) \geq m(g^*,g) = \tilde{m}(g',g)\), with the inequality strict if \(ik_j \in g^*\). If \(ik_j \notin g^*\), since \(g_1\) is unstable \((g(k_j) \text{ is single})\), Lemma 5.5 implies that there exists \(T \in \mathbb{N}_+\), \(g_T \in G\), such that \(P_0^T(g_1,g_T) > 0\) and \(\tilde{m}(g',g_T) > \tilde{m}(g',g_1) = \tilde{m}(g',g)\).

Case II: \((g(i) \neq \emptyset, K \neq K^* \text{ and } \emptyset \neq g(k_j) \in g^*(K_j) \text{ or } g(k_j) = \emptyset)\) and \((ik_j \notin g^*)\).

By definition of \(g^*\), it must be that \(g(k_j) = g^*(k_j)\). Note that \(m(g^*,g_1) \geq m(g^*,g) - 2\) as \(g_1(k_j) \neq g(k_j) = g^*(k_j)\).

If \(g^*(g(i)) = \emptyset\), then \(g(K) = g^*(K)^{23}\), so \(g^*(i) \in K\) and we have a contradiction. Therefore \(g^*(g(i)) \neq \emptyset\).

First, suppose that \(g^*(g(i))\) is indifferent between \(g\) and \(g^*\). This implies \(g(g^*(g(i)))\) and \(g^*(g^*(g(i))) = g(i)\) are positions in the same college. By definition of \(g^*\) and lemma 5.6 we have \(g(g^*(g(i))) = g^*(g^*(g(i))) = g(i)\). So \(g^*(g(i)) = g^*(k_j)\) which is a contradiction. Therefore \(g^*(g(i))\) is not indifferent between \(g\) and \(g^*\).

Second, suppose that \(g^*(g(i))\) prefers \(g^*\) to \(g\). Let \(g_2 = g_1 + g(i)g^*(g(i))\). Recall that \(g(i)\) is single, that \(g^*(g(i))\) is either single or indifferent between \(g\) and \(g_1\), and that Assumption 4 does not prevent \(g(i)\) and \(g^*(g(i))\) from being matched, so \(P_0(g_1,g_2) > 0\). Note that \(\tilde{m}(g',g_2) \geq m(g^*,g_2) \geq m(g^*,g) = \tilde{m}(g',g)\). It cannot be that \(g^*(g(i)) = g(k_j)\) as by \(g(k_j) = g^*(k_j)\) we then have that \(g^*(g(i)) = g^*(k_j)\) which would imply \(g(i) = k_j\), contradicting \(g + ik_j \in L(g)\). If \(g(k_j) \neq \emptyset\), \(g_2\) is unstable because \(g(k_j)\) is single. If \(g(k_j) = \emptyset\), then for all \(g \in \mathcal{C}\), \(g(K_j) = g(K_j)\), so \(g(i) \notin K_j\) implies \(ik_j\) is not in any stable matching and \(g_2\) is unstable. Lemma 5.5 implies that there exists \(T \in \mathbb{N}_+\), \(g_T \in G\), such that \(P_0^T(g_2,g_T) > 0\) and \(\tilde{m}(g',g_T) > \tilde{m}(g',g_2) \geq \tilde{m}(g',g)\).

Third, suppose that \(g^*(g(i))\) prefers \(g\) to \(g^*\). Corollary A.1 implies that \(g(i)\) prefers \(g^*\) to \(g\), and that \(i\) prefers \(g\) to \(g^*\), and that \(g^*(i)\) prefers \(g^*\) to \(g\).^{24} If \(k_j\) prefers \(g^*\) to \(g_1\) and \(g^*(k_j) \neq \emptyset\), then let \(k_j\) and \(g^*(k_j)\) get matched.^{25} If \(g^*(k_j) = \emptyset\), let \(k_j\) leave \(i\) to become a singleton. Let the resulting network be \(g_3\). Note that \(\tilde{m}(g',g_3) \geq m(g^*,g_3) \geq m(g^*,g) = \tilde{m}(g',g)\). Since \(g_3\) is unstable (\(i\) is single), Lemma 5.5 implies that there exists \(T \in \mathbb{N}_+\), \(g_T \in G\), such that \(P_0^T(g_3,g_T) > 0\) and \(\tilde{m}(g',g_T) > \tilde{m}(g',g_3) \geq \tilde{m}(g',g)\). If \(k_j\) prefers \(g_1\) to

---

23Roth (1986) tells us that 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.

24They strictly do so. \((g(i)\) and \(g^*(i)\) are college positions and have strict preferences over students. \(K \neq K^*\) implies that student \(i\) has strict preferences over \(g\) and \(g^*\).

25If \(g(k_j) \neq \emptyset\) then \(g^*(k_j)\) is single in \(g_1\) because \(g(k_j) = g^*(k_j)\). Assumption 4 does not prevent \(k_j\) and \(g^*(k_j)\) from being matched.
\( g^* \), then \( i \) prefers \( g^* \) to \( g_1 \). \(^{26}\) \( g^*(i) \neq k_j \), so \( g_1(g^*(i)) = g(g^*(i)) \) and \( g^*(i) \) does not prefer \( g_1 \) to \( g \). Therefore \( g^*(i) \) prefers \( g^* \) to \( g_1 \). Let \( i \) and \( g^*(i) \) get matched. Let \( g_4 \) denote the resulting network. Note that \( \bar{m}(g', g_4) \geq m(g^*, g_4) \geq m(g^*, g) = \bar{m}(g', g) \). \( g(i) \) is single and \( g^*(i) \not\in K \), so \( |g_4(K)| < |g(K)| \) and \( g_4 \) is unstable. Lemma 5.5 implies that there exists \( T \in \mathbb{N}_+ \), \( g_T \in G \), such that \( P_T^T(g_4, g_T) > 0 \) and \( \bar{m}(g', g_T) > m(g', g_4) \geq \bar{m}(g', g) \).

Case III: \( (g(i) = \emptyset \) or \( g(i) \neq \emptyset \), \( K = K^* \)) and \( (\emptyset \neq g(k_j) \not\in g^*(K_j)) \) and \( (ik_j \not\in g^*) \).

Note that \( m(g^*, g_1) \geq m(g^*, g) - 2 \) as \( g(k_j) \not\in g^*(K_j) \). 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 \( k_j \) to be single. Let \( g_5 \) denote the resulting network. \( g(i) \in K = K^* \) or \( g(i) = g^*(i) = \emptyset \) implies that \( \bar{m}(g', g_5) \geq \bar{m}(g^*, g) \). Since \( g_5 \) is unstable (\( g(k_j) \) is single), Lemma 5.5 implies that \( \bar{m}(g', g_T) > \bar{m}(g', g_5) \geq \bar{m}(g', g) \) with \( P_T^T(g_5, g_T) > 0 \) for some \( T \in \mathbb{N}_+ \), \( g_T \in G \).

Next, suppose that \( i \) prefers \( g_1 \) to \( g \). Then \( k_j \) prefers \( g \) and \( g^* \) to \( g_1 \). If \( k_j \) prefers \( g^* \) to \( g \), then \( g(k_j) \) prefers \( g \) to \( g^* \). This implies that \( g^*(g(k_j)) \) prefers \( g^* \) 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 (1986) we have \( g^*(g(k_j)) = g(k_j) = \emptyset \), contradicting our assumptions for Case III. If \( g(g^*(g(k_j))) \neq \emptyset \), then \( g^*(g(k_j))) \) is single in \( g_6 \), so \( g_6 \) is unstable. Lemma 5.5 implies that \( \bar{m}(g', g_T) > \bar{m}(g', g_6) \geq \bar{m}(g', g) \) with \( P_T^T(g_6, g_T) > 0 \) for some \( T \in \mathbb{N}_+ \), \( g_T \in G \).

If \( k_j \) prefers \( g \) to \( g^* \), then \( g^*(k_j) \) prefers \( g^* \) to \( g \). Let \( k_j \) and \( g^*(k_j) \) get matched. Let \( g_7 \) denote the resulting network. Note that \( \bar{m}(g', g_7) \geq \bar{m}(g^*, g_7) \geq \bar{m}(g^*, g) = \bar{m}(g', g) \). Since \( g_7 \) is unstable (\( g(k_j) \) is single), Lemma 5.5 implies that \( \bar{m}(g', g_T) > \bar{m}(g', g_7) \geq \bar{m}(g', g) \) with \( P_T^T(g_7, g_T) > 0 \) for some \( T \in \mathbb{N}_+ \), \( g_T \in G \).

For all cases, we have shown that there exists \( T \in \mathbb{N}_+ \), \( g_T \in G \), such that \( P_T^T(g_1, g_T) > 0 \) and \( \bar{m}(g', g_T) > \bar{m}(g', g) \). If \( g_T \in \mathcal{C} \), then we are done by letting \( g_T = g'' \). If \( g_T \not\in \mathcal{C} \), then repeated application of Lemma 5.5 will lead the process to \( g'' \in \mathcal{C} \) such that \( \bar{m}(g', g'') > \bar{m}(g', g) \).

**Proof of Theorem 5.8.** If \( g \in SS \), then \( g \in \mathcal{C} \) and there exists a minimal cost spanning tree rooted at \( g \). Denote the cost of this tree by \( cost(g) \). Assume \( g \not\in OS \). Choose \( g' \in OS \). Construct a path of edges \( (g = g^1, \ldots, g^L) \) such that \( g^i \in \mathcal{C} \), \( g^i \not\in OS \) for \( i = 1, \ldots, L - 1 \), and \( g^L \in OS \). The path is constructed as follows. For each \( g^i, i = 1, \ldots, L - 1 \), Lemma A.4

---

\(^{26}\) \( i \) strictly does so. \( ik_j \not\in g^* \) and the definition of \( g^* \) imply that \( i \) is matched to different colleges in \( g_1 \) and \( g^* \).
implies:

\[ \exists g^{i+1} \in \mathcal{C}: \bar{m}(g', g^{i+1}) > \bar{m}(g', g^i) \quad \text{and} \quad C(g^i, g^{i+1}) = c_L(g^i). \]

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 \( g^2, \ldots, g^{L-1} \). Remove the edge leaving \( g^L \). Denote the cost of the new tree by \( \text{cost}(g^L) \). Then:

\[ \text{cost}(g^L) \leq \text{cost}(g) + c_L(g) - c_L(g^L) < \text{cost}(g). \]

The first inequality follows from the construction of the tree rooted at \( g^L \); the second inequality holds as \( g \notin OS \) implies \( c_L(g) < c_L(g^L) \). So, the conjectured minimal cost spanning tree can have been no such thing. Contradiction. ■

**References**


