THE UNIVERSITY OF SYDNEY

Economics Working Paper Series

2014 - 05

Stochastic Stability in Assignment Problems

Bettina Klaus and Jonathan Newton

April 2014

Stochastic Stability in Assignment Problems*

Bettina Klaus† Jonathan Newton‡

April 3, 2014

Abstract

In a dynamic model of assignment problems, small deviations suffice to move between stable outcomes. This result is used to obtain no-selection and almost-no-selection results under the stochastic stability concept for uniform and payoff-dependent errors. There is no-selection of partner or payoff under uniform errors, nor for agents with multiple optimal partners under payoff-dependent errors. There can be selection of payoff for agents with a unique optimal partner under payoff-dependent errors. However, when every agent has a unique optimal partner, almost-no-selection is obtained.

JEL classification: C71, C78, D63.

Keywords: Assignment problem, (core) stability, decentralization, stochastic stability.

1 Introduction

We study two-sided one-to-one matching markets with side payments. Two-sided matching markets with side payments –*assignment problems*– were first analyzed by Shapley and Shubik (1971). In an assignment problem, indivisible objects (e.g., jobs) are exchanged with monetary transfers (e.g., salaries) between two finite sets of agents (e.g., workers and firms). Agents are heterogeneous in the sense that each object may have different values to different agents. Each agent either demands or supplies exactly one unit. Thus, agents form pairs to exchange the corresponding objects and at the same time make monetary transfers.

An outcome for an assignment problem specifies a matching between agents of both sides of the market and, for each agent, a payoff. An outcome is in the *core* if no coalition of agents can improve their payoffs by rematching among themselves.

This paper adds to the literature on the dynamics of assignment problems. It has been recently shown that under plausible dynamics of rematching and surplus sharing, convergence to the core of the assignment problem is assured (Chen, Fujishige, and Yang, 2012; Biró, Bomhoff, Golovach, Kern, and Paulusma, 2013; Klaus and Payot, 2013; Nax, Pradelski, and Young, 2013).

*The authors would like to thank Heinrich Nax, Bary Pradelski, Jan Christoph Schlegel, Damian Sercombe, and Markus Walzl for detailed comments and questions.

†Faculty of Business and Economics (HEC), University of Lausanne, Internef 538, CH-1015 Lausanne, Switzerland; e-mail: bettina.klaus@unil.ch

‡School of Economics, University of Sydney, NSW 2006, Australia; e-mail: jonathan.newton@sydney.edu.au

A typical such dynamic involves two agents meeting every period, and if they can improve upon their current payoffs by matching with one another, they do so. The current paper analyzes the effect of perturbations which can move the process away from core outcomes. Under such perturbations, any agent can occasionally make an error and move to an outcome which gives him a payoff lower than his current payoff. Take any core outcome and subject it to a small deviation within a single matched pair whereby one of the agents in the pair gains a unit of payoff and the other loses a unit of payoff. It is shown that such a small deviation suffices for the unperturbed blocking dynamics to subsequently move to another core outcome. More specifically, this can occur in a way that the reached optimal matching is the same as the original matching and only payoffs change (Theorem 2), or in such a way that payoffs stay the same and a different optimal matching, if one exists, is reached (Theorem 3).

A consequence of small deviations sufficing to move the process between core states is that stochastic stability, the analysis of the limit of invariant distributions of the process as perturbations become unlikely, is a weak selection concept. The identity of stochastically stable states depends on the error process. When the dynamic is perturbed by uniform errors (see Young, 1993) such that every error has the same (order of magnitude of) probability of occurring, there is no selection: every core state is stochastically stable (Theorem 5). This result is similar in spirit to the results of Jackson and Watts (2002) and Klaus, Klijn, and Walzl (2010) which find no selection in marriage and roommate problems under uniform error processes.[1]

Payoff-dependent errors occur with probabilities that depend on the payoff loss incurred when they are made. Logit errors (see Blume, 1993) are an example of such errors, and occur with probabilities that are log-linear in such payoff losses. Another possibility is that errors involving indifference occur more often than errors by which agents' payoffs strictly decrease (Serrano and Volij, 2008). We refer to these latter errors as *stepped*. Under stepped errors we find that

(i) All optimal matchings occur in some stochastically stable outcome (Theorem 6).

(ii) For any agent with different partners in different optimal matchings, any core payoff can be attained as a stochastically stable payoff (also Theorem 6).

(iii) For agents who have the same partner in every optimal matching, the set of stochastically stable payoffs can be a strict subset of the set of core payoffs (Example 1), but

(iv) if every agent has a unique optimal partner, then we obtain almost-no-selection: the interior of the set of core payoffs is stochastically stable, where the interior refers to the set of payoffs for which no two agents who are not matched to one another at the optimal matching can do at least as well by matching with one another (Theorem 7).

Finally, we define the class of *weakly payoff monotone* error processes. This class is very large. Despite this, every error process in this class is either similar to uniform errors in that Theorem 5 holds, or is similar to stepped errors in that Theorems 6 and 7 hold (Theorem 8).

[1]The marriage problem is the non-transferable utility equivalent of the assignment problem.

2 Related literature

2.1 Perturbed dynamics and selection in the core

A related literature is the literature on *convergence to the core* in cooperative games (Feldman, 1974; Green, 1974; Sengupta and Sengupta, 1996; Agastya, 1997; Serrano and Volij, 2008; Newton, 2012b). Agastya (1999) shows that if a cooperative game is modelled as a generalized Nash demand game, then the stochastically stable states are states in the core at which the maximum payoff over all agents is minimized. Newton (2012b) shows that, under some conditions, the addition of joint strategic switching to such models leads to Rawlsian selection within the strong core (referred to as the 'interior core' in the cited paper), maximizing the minimum payoff over all agents. In assignment problems, the strong core is empty as value function inequalities for matched pairs always hold with equality, so the methods of Newton (2012b) cannot be applied. Nax and Pradelski (2013) have recently shown a maxmin selection result within the core for assignment games, a result discussed next.

2.2 Nax and Pradelski (2013)

Nax and Pradelski (2013) analyze an error process in which payoffs can be shocked with a probability which is log-linear in the size of the shock. If an agent, following a shock, has a payoff lower than that which he could achieve by a change of partner, then he can change his partner. Using arguments adapted from Newton and Sawa (2013), Nax and Pradelski (2013) show that under this process the set of stochastically stable states is a subset of the least core (Maschler, Peleg, and Shapley, 1979). If any agent has multiple optimal partners, then the least core equals the core, so selection under the error processes in the current paper also select within the least core. If every agent has a unique optimal partner, then the least core can be a strict subset of the interior of the core, therefore in this case the least core inclusion of Nax and Pradelski (2013) fails under the error process of the current paper. The reason for the difference between the two papers is that the current paper allows for errors whereby the agents in a given pair remain matched yet adjust the payoffs they obtain within the pair.

2.3 Selection in matching problems

Newton and Sawa (2013) give a general selection result for matching problems (marriage problems, roommate problems, college admissions problems) for any error process, including payoff-dependent processes. All stochastically stable matchings lie within the set of matchings which are most robust to one-shot deviation. This is often a strict subset of the core. It is worth commenting on why selection is not often likewise attained for assignment problems under payoff-dependent dynamics. The reason is that in the assignment problem, there is always another core outcome in which payoffs do not differ at all, or differ only slightly, from the payoffs of the current outcome. If any agent has multiple optimal partners, then every core outcome is equally robust to one-shot deviation. If every agent has a unique optimal partner, then every interior core outcome is equally robust to one-shot deviation. Therefore, a similar inclusion holds for assignment problems as holds for matching problems: stochastically stable states are contained in the set of one-shot stable outcomes. However, the continuous, or almost continuous in the

case of discretization, nature of the core in the assignment problem means that this inclusion has less selective power.

3 The Assignment Problem

We consider a simple labor market model that matches firms and workers. Let W and F be two distinct finite sets containing $|W|$ workers and $|F|$ firms, respectively. Thus, the set of agents equals $W \cup F$. We denote generic agents by i, j, a generic worker by w, and a generic firm by f. We assume that each worker can work for at most one firm and a firm can employ at most one worker.[2] We denote the set of pairs that agents in $W \times F$ can form (including "degenerate" pairs where agents $i \in W \cup F$ form a "pair" (i, i) with themselves) by $P(W, F) = \{(w, f) \in W \times F\} \cup \{(i, i) \mid i \in W \cup F\}$.

A function $v : P(W, F) \to \mathbb{N}_0$ is a *value function* for $W \cup F$ if for each $i \in W \cup F$, $v(i, i) = 0$. The value function v describes the *value* (in nonnegative integers) that agents create when forming pairs. In particular, $v(i, i) = 0$ represents the *reservation value* of an agent $i \in W \cup F$.[3] A *(two-sided one-to-one) assignment problem* is a triple (W, F, v).

A *matching* μ (for assignment problem (W, F, v)) is a function $\mu : W \cup F \to W \cup F$ of order two (that is, $\mu(\mu(i)) = i$) such that

(i) for $w \in W$, if $\mu(w) \neq w$, then $\mu(w) \in F$ and

(ii) for $f \in F$, if $\mu(f) \neq f$, then $\mu(f) \in W$.

Two agents $i, j \in W \cup F$ *are matched* if $\mu(i) = j$ (or equivalently $\mu(j) = i$); for convenience, we also use the notation $(i, j) \in \mu$. We refer to $\mu(i)$ as i's *partner at* μ. If $(w, f) \in \mu$, then we say that worker w and firm f form a *couple at* μ. If $(i, i) \in \mu$, then we say that agent i *remains single at* μ. Thus, at any matching μ, the set of agents is partitioned into the set of agents that form couples $C(\mu) := \{i \in W \cup F \mid \mu(i) \neq i\}$ and the set of agents that remain single $S(\mu) := \{i \in W \cup F \mid \mu(i) = i\}$; i.e., $W \cup F = C(\mu) \cup S(\mu)$. Let $\mathcal{M}(W, F)$ denote the set of matchings (for W and F).

A matching μ is *optimal* (for assignment problem (W, F, v)) if, for all matchings $\mu' \in \mathcal{M}(W, F)$,

$$\sum_{(i,j)\in\mu} v(i, j) \geq \sum_{(i,j)\in\mu'} v(i, j).$$

[2] This unit-demand assumption has also been made in the following closely related articles: Shapley and Shubik (1971), Crawford and Knoer (1981), Chen et al. (2012), Biró et al. (2013), Klaus and Payot (2013), and Nax et al. (2013).

[3] It is convenient to normalize agents' reservation values to be all equal to zero, i.e., one only measures net gains from the stand alone value each agent can obtain. This normalization, for instance, can be obtained by assuming that for each $(w, f) \in W \times F$, worker w requires a minimal salary $s_{\min}(w, f)$ to work for firm f and firm f is willing to pay a maximal salary $s_{\max}(w, f)$ for worker w. Then, taking the possibility of not forming a pair into account, the joint value created equals $v(w, f) = \max\{(s_{\max}(w, f) - s_{\min}(w, f)), 0\} \geq 0$. Our assumption that values are integers is also uncontroversial by assuming that a smallest monetary unit of exchange exists.

If μ is an optimal matching, then we refer to $\mu(i)$ as i's *optimal partner at* μ. We say that a worker w and a firm f *are optimal partners* if there exists at least one optimal matching μ such that $(w, f) \in \mu$.

An *outcome* (for assignment problem (W, F, v)) is a pair $(\mu, u) \in \mathcal{M}(W, F) \times \mathbb{N}_0^{|W \cup F|}$ where μ is a matching and u is a payoff vector such that

(i) if $(w, f) \in \mu$, then $u_w + u_f = v(w, f)$, and

(ii) if $(i, i) \in \mu$, then $u_i = v(i, i) = 0$.[4]

Let $\mathcal{O}(W, F, v)$ denote the set of outcomes (for assignment problem (W, F, v)).

It is typical to refer to an outcome (μ, u) [a payoff vector u] as *individually rational* if for each $i \in W \cup F$, $u_i \geq 0$. Our assumption that payoffs are non-negative at any outcome means that this is automatically satisfied in our model. Without the restriction that payoffs be non-negative, the set of possible outcomes for a given assignment problem are countably infinite rather than finite. The results of the paper hold for either case, but for the sake of simplicity of exposition, we include the restriction in our model.

If, at outcome (μ, u) [at payoff vector u], there is a pair $(w, f) \in W \times F$ such that $u_w + u_f < v(w, f)$, then w and f have an incentive to form a couple in order to obtain a higher payoff. Then, (w, f) is a *blocking pair* for outcome (μ, u) [for payoff vector u] that creates the *blocking surplus*

$$bs(u; (w, f)) := v(w, f) - u_w - u_f > 0.$$

Throughout this article we will use weak blocking, i.e., the blocking pair divides the blocking surplus such that both agents are weakly better off and at least one of them is strictly better off. Note that when any blocking pair matches, were they to randomly allocate the blocking surplus between them, they would both strictly gain in expectation.

An outcome (μ, u) [a payoff vector u] is *(core) stable* if no blocking pairs $(i, j) \in P(W, F)$ exist, that is,

(a) for all $i \in W \cup F$, $u_i \geq 0$ and

(b) for all $(w, f) \in W \times F$, $u_w + u_f \geq v(w, f)$.

Let $\mathcal{S}(W, F, v)$ denote the *set of (core) stable outcomes* (for assignment problem (W, F, v)).

The following lemma explains that single agents always receive their reservation value at a stable outcome and how optimal matchings and stable payoffs are related.

Lemma 1 (**Stability: single agents and optimal matchings**).
(a) *If an agent is single at a stable outcome, then at each stable outcome, he receives his reservation value (Demange and Gale, 1985).*

[4]Since later on we will consider (core) stability and (weak) blocking, it is without loss of generality to use conditions (i) and (ii) to define an outcome instead of the more standard requirement that $\sum_{i \in W \cup F} u_i = \sum_{(i,j) \in \mu} v(i, j)$.

(b) *If (μ, u) is a stable outcome for assignment problem (W, F, v), then μ is an optimal matching for assignment problem (W, F, v)* (Roth and Sotomayor, 1990, Corollary 8.8).
(c) *Let (μ, u) be a stable outcome and μ' be an optimal matching for assignment problem (W, F, v). Then, (μ', u) is a stable outcome for assignment problem (W, F, v)* (Roth and Sotomayor, 1990, Corollary 8.7).

The following lemma states some facts about the payoff structure obtained for the set of stable outcomes. First, we define agents' minimal and maximal stable payoffs (which are well-defined; see Shapley and Shubik, 1971).

Let (W, F, v) be an assignment problem. Then, for each agent $i \in W \cup F$ the *set of stable payoffs* equals $[\underline{u}_i, \overline{u}_i] \cap \mathbb{N}_0 \equiv \{u'_i \in \mathbb{N}_0 \mid \underline{u}_i \leq u'_i \leq \overline{u}_i\} = \{u'_i \in \mathbb{N}_0 \mid$ there exists a stable outcome (μ, u) such that $u_i = u'_i\}$. Thus, $\underline{u}_i$ is the *minimal stable payoff of agent i* and $\overline{u}_i$ is the *maximal stable payoff of agent i*. Let $\underline{u}_W \equiv (\underline{u}_w)_{w \in W}$, $\underline{u}_F \equiv (\underline{u}_f)_{f \in F}$, $\overline{u}_W \equiv (\overline{u}_w)_{w \in W}$, and $\overline{u}_F \equiv (\overline{u}_f)_{f \in F}$. If, at some arbitrary outcome (μ, u), agent i receives a payoff $u_i \notin [\underline{u}_i, \overline{u}_i]$, then we say that agent i receives an *unstable payoff*.

Lemma 2 (**Side-optimal stable outcomes, Shapley and Shubik, 1971, Theorem 3**). *Let μ be an optimal matching for assignment problem (W, F, v). Then, outcomes $(\mu, (\overline{u}_W, \underline{u}_F))$ and $(\mu, (\underline{u}_W, \overline{u}_F))$ [payoff vectors $(\overline{u}_W, \underline{u}_F)$ and $(\underline{u}_W, \overline{u}_F)$] are stable.*

Let μ be an optimal matching for assignment problem (W, F, v). Then, outcome $(\mu, (\overline{u}_W, \underline{u}_F))$ (payoff vector $(\overline{u}_W, \underline{u}_F)$) is a *worker-optimal stable outcome* (the *worker-optimal stable payoff vector*) and outcome $(\mu, (\underline{u}_W, \overline{u}_F))$ (payoff vector $(\underline{u}_W, \overline{u}_F)$) is a *firm-optimal stable outcome* (the *firm-optimal stable payoff vector*).

4 Blocking Paths to Stability

A *path* (for assignment problem (W, F, v)) is a sequence of outcomes $(\mu^1, u^1), ..., (\mu^k, u^k)$ such that for all $l \in \{1, ..., k-1\}$, the outcome (μ^{l+1}, u^{l+1}) *is obtained from* (μ^l, u^l) by matching a pair $(i_l, j_l) \in P(W, F)$. This induces the matching μ^{l+1}

$$\mu^{l+1}(x) = \begin{cases} j_l & \text{if } x = i_l, \\ i_l & \text{if } x = j_l, \\ x & \text{if } x \neq i_l, j_l \text{ and } x \in \{\mu^l(i_l), \mu^l(j_l)\}, \\ \mu^l(x) & \text{otherwise} \end{cases}$$

and the payoff vector u^{l+1}

$$u_x^{l+1} = \begin{cases} u_{i_l}^{l+1} & \text{if } x = i_l, \\ u_{j_l}^{l+1} & \text{if } x = j_l, \\ 0 & \text{if } x \neq i_l, j_l \text{ and } x \in \{\mu^l(i_l), \mu^l(j_l)\}, \\ u_x^l & \text{otherwise} \end{cases}$$

such that $u_{i_l}^{l+1} + u_{j_l}^{l+1} = v(i_l, j_l)$ if $i_l \neq j_l$ and $u_{i_l}^{l+1} = u_{j_l}^{l+1} = 0$ otherwise. Thus, at outcome (μ^{l+1}, u^{l+1}), agents i_l and j_l are matched and generate value $v(i_l, j_l)$, their former partners

(unless i_l and j_l were matched to each other already) are single and receive zero payoffs, and all the other agents are matched to the same partners and obtain the same payoffs as before.

Let outcome $(\hat{\mu}, \hat{u})$ be obtained from outcome (μ, u) by matching the pair $(w, f) \in W \times F$. Then, we say an error has been made if $\hat{u}_w < u_w$ or $\hat{u}_f < u_f$. We say that agent w made a *1-error* if $\hat{u}_w = u_w - 1$. Similarly, we say that agent f made a *1-error* if $\hat{u}_f = u_f - 1$.

A *blocking path* (for assignment problem (W, F, v)) is a path $(\mu^1, u^1), ..., (\mu^k, u^k)$ such that for all $l \in \{1, ..., k-1\}$, the outcome (μ^{l+1}, u^{l+1}) is obtained from (μ^l, u^l) by matching a blocking pair $(w_l, f_l) \in W \times F$ for (μ^l, u^l) and their payoffs are $u^{l+1}_{w_l} \geq u^l_{w_l}$ and $u^{l+1}_{f_l} \geq u^l_{f_l}$ with at least one strict inequality, i.e., the blocking pair (w_l, f_l) splits their blocking surplus such that each of them is weakly better off and at least one of them is strictly better off at outcome (μ^{l+1}, u^{l+1}). In words, a blocking path is a finite sequence of outcomes at every step of which two agents pair up, breaking with any existing partners, and sharing surplus so that neither agent is worse off and at least one is better off. We say that a blocking path *leads to stability* if the last outcome (μ^k, u^k) is stable. Note that we are using *weak blocking* in our definition of a blocking path.

Various recent papers (Chen et al., 2012; Biró et al., 2013; Klaus and Payot, 2013; Nax et al., 2013) have proven that for any assignment problem and from any unstable outcome, a path to stability exists.

Theorem 1 (Paths to stability)**.** *Let (W, F, v) be an assignment problem and $(\mu, u) \in \mathcal{O}(W, F, v)$. Then, there exists a blocking path $(\mu, u) = (\mu^1, u^1), ..., (\mu^k, u^k)$ that leads to stability, i.e., (μ^k, u^k) is stable.*

Next, for any two payoff vectors $u, u' \in \mathbb{N}_0^{|W \cup F|}$ we define a (coordinatewise) *payoff distance vector* $d(u, u') \equiv (\mid u_i - u'_i \mid)_{i \in W \cup F}$. We say that outcome $(\bar{\mu}, \bar{u})$ is *payoff closer* to outcome (μ', u') than outcome (μ, u)) is, if and only if for all $i \in W \cup F$, $\mid u'_i - \bar{u}_i \mid \leq \mid u'_i - u_i \mid$ with strict inequality for at least one $i \in W \cup F$.

Take any assignment problem and any two stable outcomes (μ, u) and (μ', u') with different payoffs. The next result shows that, starting from (μ, u), following a single 1-error, there is a path to stability to some stable outcome $(\mu, \bar{u})$, which is payoff closer to (μ', u') than (μ, u) was. Note that outcome $(\mu, \bar{u})$ has the same matching as the initial outcome (μ, u). That is, we can attain a new stable outcome which is payoff closer to a target outcome and retain the original matching.

Theorem 2 (Moving closer I)**.** *Let (W, F, v) be an assignment problem and $(\mu, u), (\mu', u') \in \mathcal{S}(W, F, v)$ with $u \neq u'$. Then, there exists a path $(\mu, u), (\mu, \hat{u}), (\mu^1, u^1), ..., (\mu, \bar{u})$ such that*
(i) *outcome $(\mu, \hat{u})$ is obtained from (μ, u) by (re)matching a pair $(w, f) \in \mu$ such that either worker w or firm f makes a 1-error,*
(ii) *$(\mu, \hat{u}), (\mu^1, u^1), ..., (\mu^k, u^k), (\mu, \bar{u})$ is a blocking path that leads to stability, and*
(iii) *(stable) outcome $(\mu, \bar{u})$ is payoff closer to (μ', u') than (μ, u) is.*

We prove Theorem 2 in Appendix A. Loosely speaking, the proof works as follows. Let (μ, u) be the starting stable outcome and (μ', u') be the target stable outcome. Then, we first let a matched pair (w, f) change their payoff such that they make a 1-error that brings them payoff closer to u'. Assume that being payoff closer to u' requires the 1-error to be such that the

worker loses one unit of payoff and the firm gains one unit of payoff. We then show how this can trigger a blocking path where more and more worker-firm pairs are first unmatched and then rematched to receive payoffs such that the worker loses one unit of payoff and the firm gains one unit of payoff (and this is payoff closer to u'). This unmatch and rematch procedure stops at an outcome $(\mu, \bar{u})$ that is stable and payoff closer to the target stable outcome (μ', u').

Given two matchings $\mu, \mu' \in \mathcal{M}(W, F)$, let $m(\mu, \mu')$ denote the number of agents that have the same partner under μ and μ', i.e., $m(\mu, \mu') = |\{i \in N : \mu(i) = \mu'(i)\}|$. Interpreting $m(\cdot, \cdot)$ as quantifying the distance between two matchings we say outcome $(\bar{\mu}, \bar{u})$ is *match closer* to outcome (μ', u') than outcome (μ, u) is, if and only if $m(\mu', \bar{\mu}) > m(\mu', \mu)$.

In Theorem 2 we showed how to move to a stable outcome that has the same underlying optimal matching as the starting stable outcome and a payoff vector closer to that of the target stable outcome. The next result performs the opposite trick, showing that (in the absence of matched pairs which do not create value) we can move to a stable outcome that has the same payoff vector as the starting stable outcome and an optimal matching closer to that of the target stable outcome.

Theorem 3 (Moving closer II)**.** *Let (W, F, v) be an assignment problem and $(\mu, u), (\mu', u) \in \mathcal{S}(W, F, v)$ with $\mu \neq \mu'$ and u such that for all $i \neq \mu(i)$, $u_i + u_{\mu(i)} > 0$ and for all $i \neq \mu'(i)$, $u_i + u_{\mu'(i)} > 0$. Then, there exists a path $(\mu, u), (\mu, \hat{u}), (\mu^1, u^1), ..., (\bar{\mu}, u)$ such that*
(i) *outcome $(\mu, \hat{u})$ is obtained from (μ, u) by (re)matching a pair $(w, f) \in \mu$ such that either worker w or firm f makes a 1-error,*
(ii) *$(\mu, \hat{u}), (\mu^1, u^1), ..., (\bar{\mu}, u)$ is a blocking path that leads to stability, and*
(iii) *(stable) outcome $(\bar{\mu}, u)$ is match closer to (μ', u) than (μ, u) is.*

We prove Theorem 3 in Appendix A. Loosely speaking, the proof works as follows. Let (μ, u) be the starting stable outcome and (μ', u) be the target stable outcome. Take pair (w, f) which is matched in μ but not in μ'. Let (w, f) change their payoffs such that they make a 1-error such that, without loss of generality, the worker loses one unit of payoff and the firm gains one unit of payoff. We then show how this can trigger a blocking path where more and more worker-firm pairs are unmatched from their μ-partners and rematched with their partners according to the target optimal matching μ'. When agents rematch they obtain their original payoffs given by u. This unmatch and rematch procedure stops at an outcome $(\bar{\mu}, u)$ that is stable and match closer to the target stable outcome (μ', u). Note that agents which have the same partner in every optimal matching are never required to make errors as part of this proof. This fact is later used in the proof of Theorem 6.

Theorem 2 and Theorem 3 will help us to prove results on stochastic stability, but it is also of independent interest that only small adjustments to the payoff shares of a matched pair are necessary to move across the set of stable outcomes, and that payoffs and matchings can be adjusted independently of one another.

5 Stochastic Stability

5.1 The Unperturbed Blocking Dynamics and Absorbing Outcomes

For each assignment problem (W, F, v), we model the so-called unperturbed blocking dynamics by a *Markov process* $(\mathcal{O}, T)$, where the *state space* is the set of outcomes $\mathcal{O} = \mathcal{O}(W, F, v)$ and T is a *transition matrix* that induces the following *unperturbed blocking dynamics*. First, we shall need some notation. By $o \in \mathcal{O}$, $o = (\mu, u)$, we will denote a representative outcome.

Next, define the set of outcomes that can be obtained from outcome o by matching $(i, j) \in P(W, F)$, no matter how value is shared by i and j after they match, by $A(o, i, j) := \{o' \in \mathcal{O} \mid o' \text{ is obtained from } o \text{ by matching } (i, j)\}$. Recall that possibly $i = j$ (in which case $u_i = 0$).

Then, denote by $B(o, i, j) \subset A(o, i, j)$ the set of outcomes obtainable from outcome o via weak blocking by $(i, j) \in P(W, F)$. Given our assumption of nonnegative payoffs, a single agent can never weakly block, so if $i = j$, then $B(o, i, j) = \emptyset$. When $i \neq j$, payoffs of outcomes in $B(o, i, j)$ are such that i and j are weakly better off than they are at outcome o, and at least one member of the blocking pair is strictly better off. $B(o, i, j) := \{o' = (\mu', u') \in A(o, i, j) \mid u'_i \geq u_i,\ u'_j \geq u_j, \text{ and } u'_i + u'_j > u_i + u_j\}$. Note that if (i, j) is not a blocking pair for outcome o, then $B(o, i, j)$ is empty.

In each period $t = 1, 2, \ldots$, the process is at an outcome $o^t = (\mu^t, u^t) \in \mathcal{O}$. A pair $(i, j) \in P(W, F)$ of agents (possibly $i = j$) is randomly selected from a distribution $G(.)$ with probability mass function $g(.)$ and full support on $P(W, F)$. Let o' be chosen randomly from a distribution $H_{A(o^t,i,j)}(.)$ with probability mass function $h_{A(o^t,i,j)}(.)$ and full support on $A(o^t, i, j)$. Let

$$o^{t+1} = \begin{cases} o' & \text{if } o' \in B(o^t, i, j), \\ o^t & \text{otherwise.} \end{cases}$$

Note that $o^{t+1} \neq o^t$ implies that o^{t+1} is obtained from outcome o^t via weak blocking of a blocking pair $(i, j) \in W \times F$. If $o' \notin B(o^t, i, j)$, then $o^{t+1} = o^t$. If (i, j) is not a blocking pair for outcome o^t, then it must be that $o' \notin B(o^t, i, j)$ as $B(o^t, i, j)$ is the empty set. The dynamics as defined above will always follow a blocking path. Moreover, starting from any outcome o, as any $(i, j) \in P(W, F)$ has positive probability of being chosen, and has positive probability of moving the process to any outcome in $B(o, i, j)$, it must be that any (finite) blocking path starting from o has positive probability of being followed by the dynamics.

For two outcomes $o, o' \in \mathcal{O}$, let $T(o, o')$ denote the probability that the process moves from outcome o to o' from one period to the next. Similarly, let $T^l(o, o')$ denote the l-period transition probability, the probability that $o^{t+l} = o'$ conditional on $o^t = o$. Note that for two outcomes $o, o' \in \mathcal{O}$, $o \neq o'$ and $T(o, o') > 0$ if and only if outcome o' is obtained from o via weak blocking. Similarly, for each $l \in \mathbb{N}$, $T^l(o, o') > 0$ if and only if there exists a blocking path of at most length l from o to o'. For a set of outcomes $O \subseteq \mathcal{O}$, define $T^l(o, O) := \sum_{o' \in O} T^l(o, o')$. Note that for any $o \in \mathcal{O}$, $T(o, \mathcal{O}) = 1$.

Note that blocking pairs who weakly block are always better off in the short run (even though they might be worse off later). That is, agents are myopic but they do not make mistakes. We will consider a dynamic process with a positive probability of errors (or mistakes, or perturbations)

in Section 5.2. For further reference we therefore label the blocking process as defined in this section as the *unperturbed blocking dynamics*. The following theorem, which corresponds to Theorem 1 of Nax et al. (2013), shows that (i) every stable outcome is an absorbing state of the unperturbed blocking dynamics, and that (ii) from any starting point, the unperturbed blocking dynamics will converge to one of the stable outcomes in finite time with probability 1.

Theorem 4. *Let (W, F, v) be an assignment problem. Then,*
(i) *for all $o \in \mathcal{S}(W, F, v)$, $T(o, o) = 1$ and*
(ii) *for all $o \in \mathcal{O}(W, F, v)$, $T^l(o, \mathcal{S}) \to 1$ as $l \to \infty$.*

The proof is simple (see Appendix A). Part (i) holds as by the definition of a stable outcome, there are no blocking pairs for any $o \in \mathcal{S}$, so o must be an absorbing state. For part (ii), Theorem 1 shows that from any $o \in \mathcal{O}$, there exists a blocking path to a stable outcome. Under the unperturbed blocking dynamics, these paths occur with positive probability. Since there are a finite number of states, the probability of such a path being followed is bounded below uniformly for all states. Therefore such a path will eventually be followed and the process will end up at a stable outcome. Note that Theorem 4 holds independently of the distributions $G(.)$ and $H_{A(.,.,.)}(.)$, as long as the full support assumptions are satisfied.

5.2 The Perturbed Blocking Dynamics

The perturbed blocking dynamics is identical to the unperturbed blocking dynamics except that there is some probability of state transitions which are not based on weak blocking: following the selection of o' by $H_{A(o^t,i,j)}(.)$, let

$$o^{t+1} = \begin{cases} o' & \text{with probability } \varepsilon^{c_{(i,j)}(o^t,o')}, \\ o^t & \text{otherwise.} \end{cases}$$

The cost function $c_{(i,j)}(o^t, o')$ takes values on $\mathbb{R}_+$ and measures the relative decay of the probabilities of various transitions in terms of an 'error' parameter ε. The more rare a transition is, the higher its cost. The cost function is set to zero for transitions that can occur under the unperturbed blocking dynamics. As $\varepsilon^0 = 1$, the probability of transitions caused by weak blocking does not decay as $\varepsilon \to 0$. The cost is positive for transitions which cannot occur under the unperturbed blocking dynamics. We refer to such transitions as *errors*. If a transition from o to o' is induced by $(i, j) \in P(W, F)$ and $c_{(i,j)}(o, o') > 0$, then we say that (i, j) has made an error. More specifically, $(i, j) \in P(W, F)$ makes an error if, following the transition, the payoffs of all $k \in \{i, j\}$ are not weakly higher with at least some $k \in \{i, j\}$ having a strictly higher payoff. Note that the process for $\varepsilon = 0$ is the unperturbed blocking dynamics. Let $\widetilde{T}^l(.,.)$, $l \in \mathbb{N}_+$, denote the transition probabilities associated with the perturbed blocking dynamics.

Note that for $\varepsilon > 0$, the perturbed blocking dynamics is irreducible and ergodic and therefore has a unique stationary distribution $\pi_\varepsilon(.)$. By well known arguments (see Young, 1998), as $\varepsilon \to 0$, the limiting distribution $\pi_\varepsilon(.) \to \pi(.)$ exists and places all probability mass on recurrent classes of the process with $\varepsilon = 0$. We know from Theorem 4 that these must be stable outcomes of the unperturbed blocking dynamics. The set of outcomes with positive measure under $\pi(.)$ is important, as for small enough perturbations, on a long enough timescale, the perturbed

blocking dynamics will be found at such outcomes with a probability close to 1. Therefore, the identity of these *stochastically stable* outcomes is important to understand the long run behavior of the perturbed blocking dynamics. The *stochastically stable outcomes* are

$$\mathcal{SS}(W, F, v, c) := \{o \in \mathcal{O} \mid \pi(o) > 0\}.$$

The identity of the stochastically stable outcomes can be expected to, and indeed does, depend on the functions $c_{(.,.)}(.,.)$.[5] Therefore it is crucial that $c_{(.,.)}(.,.)$ is such that error probabilities are plausible. Two error specifications are very common in the stochastic stability literature. Under uniform error specifications (Young, 1993), agents make mistakes and take payoff reducing actions with some uniform probability. Under logit error specifications (Blume, 1993), payoff reducing actions by an agent occur with a probability that is log-linear in his loss of payoff from taking the action in question.[6] To begin, we shall analyze the process with uniform errors.

Definition 1. An error process is uniform if

$$c_{(i,j)}(o, o') = \begin{cases} 0 & \text{if } o' = o \text{ or } o' \in B(o, i, j), \\ 1 & \text{otherwise.} \end{cases}$$

Under a uniform error process, any error occurs with the same (order of ε) probability. The following theorem shows that when errors are uniform, the set of stochastically stable outcomes and the set of stable outcomes coincide. Stochastic stability does not provide any further selection beyond that already provided by the stability concept. Moreover, the proof of the theorem only relies on two types of errors: 1-errors and errors where agents earning zero either match or unmatch amongst themselves. Therefore, errors which cause a large payoff loss to the agents who make them are not necessary to obtain this no-selection result: the entire set of stable states is traversed by low payoff-loss mistakes.

Theorem 5. *If the error process is uniform, then* $\mathcal{SS}(W, F, v, c) = \mathcal{S}(W, F, v)$.

To prove the theorem some more notation is needed. Let the set of outcomes reachable in a single step from o be denoted by

$$A(o) := \bigcup_{(i,j) \in P(W,F)} A(o, i, j).$$

If there are multiple ways of moving from o to o' in a single step, we are interested in the lowest cost way of doing so. Note that there will only ever be multiple ways of moving from o to o' in a single step if the transition involves a pair which is matched in o becoming separated in o'. The cost of separation will in general be different depending on which of the agents initiates the

[5] See Bergin and Lipman (1996) and van Damme and Weibull (2002) for more on how the orders of error probabilities have an effect on selection in perturbed adaptive dynamic models.

[6] Several papers have recently argued that coalitional behavior should be considered in models of perturbed adaptive dynamics. Newton (2012a) incorporates coalitional behavior into error processes, Newton (2012b) incorporates coalitional behavior into the unperturbed dynamic, and Sawa (2013) does both. Coalitional behavior is already an integral part of matching dynamics, as pairs constitute coalitions of size two.

separation. Under uniform errors, the cost will be the same, but this will not necessarily hold for other error specifications. With this in mind, define

$$c(o,o') = \begin{cases} \min\limits_{\substack{(i,j)\in P(W,F):\\ o'\in A(o,i,j)}} c_{(i,j)}(o,o') & \text{if } o'\in A(o), \\ \infty & \text{otherwise.} \end{cases}$$

In order to determine stochastically stable outcomes, we will also be interested in the *overall cost* of moving between any two states. In a same way that the cost function measures the rarity of one period transitions between states, overall cost measures the rarity of transitions between two states over any number of periods. Let $\mathcal{P}(o,o')$ be the set of finite sequences of outcomes $\{o^1,o^2,\ldots,o^T\}$ such that $o^1=o$, $o^T=o'$ and for $t=1,\ldots,T-1$, $o^{t+1}\in A(o^t)$. Define and denote the overall cost of moving from o to o' by

$$C(o,o') := \min_{\{o^1,\ldots,o^T\}\in\mathcal{P}(o,o')} \sum_{t=1}^{T-1} c(o^t,o^{t+1}).$$

With the concept of overall cost in hand, we can use the classic characterization results of Freidlin and Wentzell (1984) and Young (1993). A o-tree is a directed graph on $\mathcal{S}(W,T,v)$ such that every vertex except for o has outdegree 1 and the graph has no cycles. For two outcomes $o',o''\in\mathcal{O}$, $o'\to o''$ denotes the directed edge from o' to o''. Let $\mathcal{G}(o)$ denote the set of all o-trees. For $g\in\mathcal{G}(o)$, define

$$\mathcal{V}(g) := \sum_{(o'\to o'')\in g} C(o',o'') \text{ and } \mathcal{V}_{min}(o) := \min_{g\in\mathcal{G}(o)} \mathcal{V}(g).$$

That is, $\mathcal{V}(g)$ is the sum of the overall costs of all the edges in the tree g, and $\mathcal{V}_{min}(o)$ is the total cost of the least cost o-tree. Define the set of outcomes at which least cost o-trees are rooted by

$$\mathcal{L}_{min} = \{o\in\mathcal{S}(W,F,v) \mid o\in \arg\min_{o\in\mathcal{S}(W,F,v)} \mathcal{V}_{min}(o)\}.$$

We know from Freidlin and Wentzell (1984) and Young (1993) that an outcome is stochastically stable if and only if it is associated with a least cost o-tree:

$$o\in\mathcal{SS}(W,F,v,c) \Leftrightarrow o\in\mathcal{L}_{min}.$$

Theorem 5 can now be proved. The argument relies on tree pruning. Starting from any o-tree rooted at a stochastically stable outcome o, given any stable outcome $\tilde{o}$, a new tree is constructed by adding and deleting edges from the o-tree, so that it becomes a $\tilde{o}$-tree. Using the results of Theorems 2 and 3 on the unperturbed blocking dynamics, the $\tilde{o}$-tree is constructed in such a way that $\mathcal{V}_{min}(\tilde{o})\leq\mathcal{V}_{min}(o)$. This means that if $o\in\mathcal{L}_{min}$ then it must be that $\tilde{o}$ is also in $\mathcal{L}_{min}$. That is, if o is stochastically stable, $\tilde{o}$ must also be. As this holds for any $\tilde{o}\in\mathcal{S}(W,F,v)$, every state in $\mathcal{S}(W,F,v)$ must be stochastically stable. The details of this proof are given in Appendix A.

Having shown a no-selection result for uniform errors, we move to payoff-dependent errors. This paper uses weak blocking dynamics, so if two agents match such that each has the same payoff as in the previous period then this is an error and not part of the unperturbed blocking dynamic. We analyze the case for which there is a distinction between errors which cause payoff loss to the erring players, and errors which do not. The latter are referred to by Serrano and Volij (2008) as 'indifference-based coalitional mistakes'. Later, we shall see that results derived for this formulation easily extend to a large class of error processes.

Definition 2. An error process is stepped if, for all $o = (\mu, u), o' = (\mu', u') \in A(o)$,

$$c_{(i,j)}(o, o') = \begin{cases} 0 & \text{if } o' = o \text{ or } o' \in B(o, i, j), \\ 1 & \max_{k \in \{i,j\}}(u_k - u'_k) > 0, \\ \delta, \quad 0 < \delta < 1 & \text{otherwise.} \end{cases}$$

The question arises as to whether under stepped errors it is still the case that $\mathcal{SS}(W, F, v, c) = \mathcal{S}(W, F, v)$. The answer is no. Although trees can still be constructed using 1-errors, there now exists an error which is lower cost than a 1-error: precisely those errors which lead to zero payoff loss for the erring agents. We refer to such errors as *0-errors*. These errors have a cost of δ. It turns out that for agents who have different partners in different optimal matchings, 0-errors suffice to move the process to any stable payoff. However, for pairs of agents which remain matched in every optimal matching, it may be the case that 0-errors suffice to move from some stable payoffs but not from others. This is illustrated in the following example.

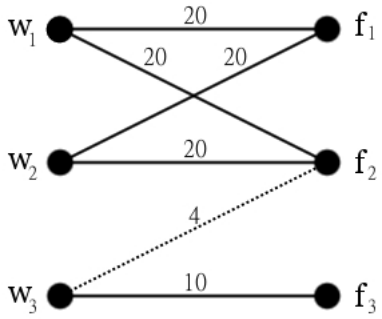

Figure 1: A line between agents w_i, f_j indicates that $v(w_i, f_j) > 0$, with the value given above the line. Solid lines indicate pairings that can arise in some optimal matching.

Example 1 (Illustrated in Figure 1)**.** Let $W = \{w_1, w_2, w_3\}$, $F = \{f_1, f_2, f_3\}$, and value function v is such that for $i, j \in \{1, 2\}$, $v(w_i, f_j) = 20$, $v(w_1, f_3) = v(w_2, f_3) = 0$, $v(w_3, f_1) = 0$, $v(w_3, f_2) = 4$, and $v(w_3, f_3) = 10$. Then $\underline{u}_W = \underline{u}_F = (0, 0, 0)$, $\bar{u}_W = \bar{u}_F = (20, 20, 10)$. At any optimal matching μ, w_1, w_2, f_1, f_2 are matched amongst themselves and w_3, f_3 are matched. Furthermore, at any stable outcome (μ, u), payoffs u are such that for $i, j \in \{1, 2\}$, $u_{w_i} + u_{f_j} = 20$, $u_{w_3} + u_{f_3} = 10$, and $u_{w_3} + u_{f_2} \geq 4$.

From a stable outcome (μ, u), if $u_{f_2} = 0$ there is a 0-error in which f_2 becomes single. If $u_{f_2} > 0$, there is a 0-error in which f_2 matches with $i \in \{w_1, w_2\}$, $i \neq \mu(f_2)$, following which

there is a weak blocking possible between i and $\mu(i) = f_1$. In either case, f_2 is left single. If $u_{w_3} < 4$ then w_3 and f_2 can block. Following this, $\mu(f_2)$ and f_2 can block. So w_3 and f_3 are left single and can rematch at any allocation of surplus such that $u_{w_3} + u_{f_3} = 10$. w_1, w_2, f_1, f_2 can then rematch amongst themselves at payoffs such that the new outcome is stable. 0-errors are all that is required to move to different stable payoffs for w_3, f_3. This is because the partnership of w_3 and f_3 can be disrupted by the activities of the other agents, even though neither w_3 nor f_3 is linked to any of the other agents in any optimal matching. However, if $u_{w_3} \geq 4$, even when f_2 is single there is no blocking possible between w_3 and f_2. Without either w_3 or f_3 making an error, there is no way in which the payoffs of these agents can change. The lowest cost such error is a 1-error.

So from any stable outcome in which $u_{w_3} < 4$, a stable outcome in which $u_{w_3} \geq 4$ can be reached with cost δ. From any stable outcome in which $u_{w_3} \geq 4$, any stable outcome in which u_{w_3} takes a different value can only be reached with cost at least 1. Therefore $(\mu, u) \in \mathcal{SS}(W, F, v, c)$ implies $u_{w_3} \geq 4$. In fact, $\mathcal{SS}(W, F, v, c)$ is precisely the set of outcomes in $\mathcal{S}(W, F, v)$ for which $u_{w_3} \geq 4$. This follows from the fact that from any of these outcomes any different stable payoffs and matching for w_1, w_2, f_1, f_2 can be reached via a single 0-error, and any different stable payoffs for w_3, f_3 can be reached via a single 1-error. □

The intuition behind Example 1 is that the bargaining position of w_3 is improved due to the latent outside option provided by a potential pairing with f_2. Although the constraint $v(w_3, f_2) \geq 4$ is not binding at any core outcome, it becomes relevant at intermediate matchings between core outcomes, facilitating transitions which increase the payoff of w_3 from values below 4. Another insight that is gained from considering Example 1 is that if 0-errors were costless, then core convergence would no longer apply in cases where there exist multiple optimal matchings.

We define the set of agents who have different partners at some stable outcomes by

$$\Delta_0 := \{i \in N \mid \text{there exist } (\mu, u), (\mu', u') \in \mathcal{S}(W, F, v) \text{ such that } \mu(i) \neq \mu'(i))\}.$$

The following lemma shows that for any given agent who has different partners at some stable outcomes and who earns a strictly positive payoff at the current outcome, we can replicate a 1-error with a 0-error. That is, for $i \in \Delta_0$ with $u_i > 0$, there exists a 0-error and a subsequent sequence of costless transitions such that the outcome at the end of the sequence is that which could have been achieved had agent i made a 1-error at the initial outcome. Note that $u_i > 0$ implies that $\mu(i) \neq i$.

Lemma 3. *Let $(\mu, u) \in \mathcal{S}(W, F, v)$ be such that for all $j \neq \mu(j)$, $u_j + u_{\mu(j)} > 0$. Let $i \in \Delta_0$, $u_i > 0$. Let (μ, u') be such that $u'_j = u_j$ for all $j \notin \{i, \mu(i)\}$, $u'_i = u_i - 1$, $u'_{\mu(i)} = u_{\mu(i)} + 1$. Then $C((\mu, u), (\mu, u')) = \delta$.*

We illustrate the proof of Lemma 3 with the example in Figure 1. Assume some stable outcome $(\mu, u) \in \mathcal{S}(W, F, v)$ such that $\mu(w_1) = f_1$, $\mu(w_2) = f_2$, and $u_{w_1} > 0$. As w_2 and f_1 are partners in some other optimal matching, it must be the case that $u_{w_2} + u_{f_1} = v(w_2, f_1)$. Then, there exists a 0-error whereby w_2 and f_1 leave their current partners and match at the same payoffs as they currently obtain. Following this, w_1 and f_2 are left single and obtain zero payoffs. For this outcome, w_1 and f_1 are a blocking pair and can costlessly rematch to obtain

payoffs $u_{w_1} - 1$ and $u_{f_1} + 1$ respectively. w_2 and f_2 are now single and can rematch at their original payoffs. The process is now at the outcome which would have been obtained had w_1 made a 1-error from the initial outcome (μ, u). A 1-error by w_1 has been replicated by a 0-error.

Lemma 3 shows that for agents who have multiple partners at some optimal matchings, 1-errors can be replicated by 0-errors. Theorem 3 shows that 1-errors suffice to move between different optimal matchings. Recall that the proof of the theorem does not require errors by $i \notin \Delta_0$, who remain with the same partner. Therefore, any optimal matching can be reached via transitions between outcomes in $\mathcal{S}(W, F, v)$ which each involve only a single 0-error. Theorem 2 shows that 1-errors suffice to move between different stable payoff vectors. Replicating these errors by 0-errors for $i \in \Delta_0$, we see that any stable payoffs for $i \in \Delta_0$ can be reached via transitions between outcomes in $\mathcal{S}(W, F, v)$ which each involve only a single 0-error. The existence of these paths of transition implies that if the initial outcome is the root of a least cost o-tree and thus stochastically stable, then the outcome reached by these paths is also the root of a least cost o-tree, and is therefore also stochastically stable. In summary, given any stable outcome o, there exists a stochastically stable outcome at which any agent with multiple optimal partners has the same partner and payoff as in outcome o.

Theorem 6. *If the error process is stepped, then for all $(\tilde{\mu}, \tilde{u}) \in \mathcal{S}(W, F, v)$, there exists $(\tilde{\mu}, u^*) \in \mathcal{SS}(W, F, v, c)$ such that for all $i \in \Delta_0$, $u_i^* = \tilde{u}_i$.*

An immediate consequence of Theorem 6 is that if every agent either has differing partners in some optimal matchings, or is single in every optimal matching, then the entire set of stable outcomes can be traversed by 0-errors, and the entire set of stable outcomes is stochastically stable. On the other hand, if there are agents who have a unique partner in all optimal matchings, then stochastic stability may or may not select a strict subset of the set of stable outcomes. In Example 1, we have $\mathcal{SS}(W, F, v, c) \subsetneq \mathcal{S}(W, F, v)$. However, if we alter the example by letting $v(w_3, f_3) = 3$, then $\mathcal{SS}(W, F, v', c) = \mathcal{S}(W, F, v')$.

Defining the set of agents who are single in every optimal matching

$$\Gamma := \{i \in W \cup F : \ \mu(i) = i \ \text{ for all } \ (\mu, u) \in \mathcal{S}(W, F, v)\}$$

we can state the following no-selection result.

Corollary 1. *If $\Delta_0 \cup \Gamma = W \cup F$, then $\mathcal{SS}(W, F, v, c) = \mathcal{S}(W, F, v)$.*

Now, consider the case that $\Delta_0 = \emptyset$. That is, there is a unique optimal matching μ and every agent has the same partner at any stable outcome. Consider two stable outcomes $o = (\mu, u), o' = (\mu, u') \in \mathcal{S}(W, F, v)$. Let (μ, u) be such that there exists a 0-error. Let (μ, u') be such that there exists no 0-error. The existence of a 0-error for (μ, u) implies that there exists $(i, j) \in P(W, F)$ such that $\mu(i) \neq j$ and $\sum_{k \in \{i,j\}} u_k = v(i, j)$. It must be that $u_i' > u_i$ and/or $u_j' > u_j$, or the same 0-error would exist from outcome (μ, u'). Assume $u_i' > u_i$. From (μ, u), following a 0-error by (i, j), let i rematch with $\mu(i)$ at payoffs $u_i + 1$, $u_{\mu_i} - 1$. Then, if $\mu(j) \neq j$, let j rematch with $\mu(j)$. The resultant outcome is the outcome that would have been obtained from (μ, u) had $\mu(i)$ made a 1-error. The resultant matching is also payoff-closer to (μ, u') than is (μ, u). Therefore by Theorem 2 there exists a costless path of the dynamic to some outcome

in $\mathcal{S}(W, F, v)$. In this manner, an o'-tree can be constructed for which any edge exiting a given outcome has the least cost of any such possible edge, using 0-errors where possible, and otherwise using 1-errors. Any tree rooted at an outcome such as o must have a higher cost as the least cost of any edge exiting o is δ, whereas the least cost of any edge exiting o' would be 1. Therefore, an outcome is stochastically stable if and only if there is no 0-error possible. We have obtained an almost-no-selection result. Let $Y \subset \mathcal{S}$ denote the *interior* of the set of stable states. That is, a stable outcome is in Y if and only if the value function inequality holds strictly for all pairs of agents who are not matched to one another and every agent's payoff is individually rational.

$$Y := \{(\mu, u) \in \mathcal{S}(W, F, v) \mid (i, j) \in P(W, F) \text{ and } \mu(i) \neq j \text{ implies } \sum_{k \in \{i,j\}} u_k > v(i, j)\}.$$

Theorem 7. *If the error process is stepped, $\Delta_0 = \emptyset$ and $Y \neq \emptyset$, then $\mathcal{SS}(W, F, v, c) = Y$.*

Note that for the non-discretized problem, a unique optimal matching implies that every agent on at least one side of the market has multiple possible core stable payoffs and that the set of stable outcomes has dimension equal to the number of agents on that side of the market (Núñez and Rafels, 2008). Therefore for a discretization which is fine enough relative to the value function, $\Delta_0 = \emptyset$ implies that $Y \neq \emptyset$ and the second condition in the statement of Theorem 7 is redundant.

Example 2. Consider Example 1 amended so that $v(w_1, f_2) = 18$. After this change there is a unique optimal matching in which w_1 is matched to f_1, w_2 to f_2, and w_3 to f_3. Every stable outcome must now satisfy $0 \leq u_{f_1} - u_{f_2} \leq 2$. If a stable outcome is such that $u_{f_1} - u_{f_2} = 0$, then by substitution $u_{f_1} - (20 - u_{w_2}) = 0$, giving $u_{f_1} + u_{w_2} = 20$. That is, the outcome cannot be in Y, as the value function constraint for f_1 and w_2 holds with equality. If a stable outcome is such that $u_{f_1} - u_{f_2} = 2$, then by substitution $(20 - u_{w_1}) - u_{f_2} = 2$, giving $u_{f_2} + u_{w_1} = 18$. That is, the outcome cannot be in Y, as the value function constraint for f_2 and w_1 holds with equality. Therefore, at any outcome in Y, and therefore at any stochastically stable outcome, it must be that $u_{f_1} - u_{f_2} = 1$. □

5.3 Generalization

The results of the paper have only depended on two types of errors, namely 0-errors and 1-errors. This fact can be used to state results for a much broader class of error processes than those considered so far.

Definition 3. An error process is weakly payoff monotone if, for all $o = (\mu, u), o' = (\mu', u') \in A(o)$,

$$c_{(i,j)}(o, o') = \begin{cases} 0 & \text{if } o' = o \text{ or } o' \in B(o, i, j), \\ & \text{otherwise,} \\ g_1(\,(u_i - u'_i)_+\,) & \text{if } i = j, \\ g_2(\,(u_i - u'_i)_+\,,\,(u_j - u'_j)_+\,) & \text{if } i \neq j. \end{cases}$$

where $g_1 : \mathbb{R}_+ \to \mathbb{R}_{++}$, $g_2 : \mathbb{R}^2_+ \to \mathbb{R}_{++}$ are non-decreasing, g_2 is symmetric in its arguments, $g_1(0) = g_2(0, 0)$, and $g_1(1) = g_2(1, 0)$.

That is, the transition cost of a mistake is a non-decreasing function of payoff losses incurred by the agents (or agent) who make the mistake. Also, the transition cost of a mistake in which no agent loses payoff or only a single agent loses a unit of payoff is the same whether the mistake is make by an agent on his own or as part of a pair.

For this class of errors, the characterizations of the previous sections apply. The reason for this is that none of the arguments so far in the paper have relied on any perturbations other than 0-errors and 1-errors. Moreover, any step between stable outcomes in our arguments only requires a single error. Therefore, if other errors have cost at least as high as that of 0-errors and 1-errors, they can be ignored for the purposes of determining stochastic stability. Which prior theorems apply depends on whether 0-errors have equal cost to, or strictly lower cost than, 1-errors.

Theorem 8. *For any weakly payoff monotone error process,*

(i) if $g_1(0) = g_1(1)$, then Theorem 5 holds, and

(ii) if $g_1(0) < g_1(1)$, then Theorems 6 and 7 hold.

Under logit errors, the cost of errors is proportional to payoff loss. To specify the cost of logit errors by pairs, the specification of Sawa (2013) can be used: the cost of a transition in which both agents in a pair lose payoff is equal to the sum of the losses.[7] To ensure strict positivity of mistake costs, δ must be added to the cost of every transition which is not part of the unperturbed blocking dynamics. This is equivalent to a formulation where a perturbation creates the opportunity for a mistake, before the actions themselves are determined by the logit choice rule.

Definition 4. An error process is logit if, for all $o = (\mu, u), o' = (\mu', u') \in A(o)$,

$$c_{(i,j)}(o, o') = \begin{cases} 0 & \text{if } o' = o \text{ or } o' \in B(o, i, j), \\ \delta + \sum_{k \in \{i,j\}} \max\{u_k - u'_k, 0\},\ \delta > 0 & \text{otherwise.} \end{cases}$$

It can immediately be seen that this rule falls into category (ii) of Theorem 8, and therefore Theorems 6 and 7 hold.

6 Conclusion

This paper makes two distinct contributions. Firstly, it improves our knowledge of paths to stability in assignment games, demonstrating that, following a small perturbation from any stable outcome, there exists a path to stability that takes the process closer to some target stable outcome. Moreover, this can be done in such a way that payoffs change and the matching remains the same (Theorem 2), or in such a way that the matching changes and payoffs remain the same (Theorem 3). The second contribution of the paper is to use these results to derive

[7] This is equivalent to saying that if one agent accepts a payoff reducing rematching with probability ε^{l_1} and the other agent accepts it with probability ε^{l_2}, then the probability of the rematching occurring is $\varepsilon^{l_1+l_2}$.

stochastic stability results for a variety of perturbed blocking dynamics. Processes with uniform errors (Theorem 5) and with two types of errors (Theorems 6 and 7) are analyzed, and a large class of perturbed processes shown to reduce to the two aforementioned cases (Theorem 8).

This paper joins a set of recent papers that have made significant progress in understanding dynamic recontracting in assignment games. However, it is still the case that such processes are understood less well than their NTU equivalents, a research area that has itself made considerable recent progress. A possible area for future work would be the study of adaptive dynamics in many-to-one and many-to-many bipartite trading networks, seeking to establish TU analogues of recent results in the NTU literature.

A Appendix

Proof of Theorem 2. Let (W, F, v) be an assignment problem and $(\mu, u), (\mu', u') \in \mathcal{S}(W, F, v)$ with $u \neq u'$. Hence, there exists $i \in W \cup F$ with $u_i \neq u'_i$. Without loss of generality, assume that for some $w \in W$, $u_w > u'_w$. By Lemma 1, $\mu(w) = f \in F$ and $u_w + u_f = u'_w + u'_f = v(w, f)$. Hence, $u_f < u'_f$.

Obtain outcome $(\mu, \hat{u})$ from (μ, u) by (re)matching agents w, f such that $\hat{u}_w = u_w - 1$ and $\hat{u}_f = u_f + 1$. Hence, agents w and f stay matched, agent w makes a 1-error, and outcome $(\mu, \hat{u})$ is payoff closer to (μ', u') than (μ, u) is. If $(\mu, \hat{u})$ is stable, then we are done.

Assume that $(\mu, \hat{u})$ is not stable. Note that outcome $(\mu, \hat{u})$ is individually rational. Hence, there exists a blocking pair $(\hat{w}, \hat{f}) \in W \times F$ with

$$\hat{u}_{\hat{w}} + \hat{u}_{\hat{f}} < v(\hat{w}, \hat{f}). \tag{1}$$

Define $T^1 \equiv \{w, f\}$.

Assume $\hat{w} \neq w$ and recall that $\hat{u}_f = u_f + 1 > u_f$ and for $i \in (W \cup F) \setminus T^1$, $\hat{u}_i = u_i$. Then, (1) implies $u_{\hat{w}} + u_{\hat{f}} < v(\hat{w}, \hat{f})$, contradicting the stability of outcome (μ, u). Hence, $\hat{w} = w$ and $\hat{f} \notin T^1$.

Since (μ, u) is stable,

$$u_w + u_{\hat{f}} \geq v(w, \hat{f}). \tag{2}$$

Using $\hat{u}_w = u_w - 1$ and $\hat{u}_{\hat{f}} = u_{\hat{f}}$, inequality (1) reads as

$$u_w + u_{\hat{f}} - 1 < v(w, \hat{f}). \tag{3}$$

Hence, inequalities (2) and (3) together imply that

$$u_w + u_{\hat{f}} = v(w, \hat{f}). \tag{4}$$

Now inequalities (1) and (4) together imply that pair $(w, \hat{f})$ can block outcome $(\mu, \hat{u})$ with payoffs $u^1_w = \hat{u}_w = u_w - 1$ and $u^1_{\hat{f}} = \hat{u}_{\hat{f}} + 1 = u_{\hat{f}} + 1$. We obtain outcome (μ^1, u^1) from outcome $(\mu, \hat{u})$ by matching blocking pair $(w, \hat{f})$ with payoffs $u^1_w \geq \hat{u}_w$ and $u^1_{\hat{f}} > \hat{u}_{\hat{f}}$.

Assume that $\mu(\hat{f}) = \hat{f}$. Then, by Lemma 1, $u_{\hat{f}} = u'_{\hat{f}} = 0$. Thus, by equation (4), $u_w = v(w, \hat{f})$. Then, our assumption that $u_w > u'_w$ implies $v(w, \hat{f}) = u_w > u'_w = u'_w + u'_{\hat{f}}$, contradicting the stability of (μ, u'). Hence, $\mu(\hat{f}) \neq \hat{f}$ and $\mu(\hat{f}) \in W$. Let $\tilde{w} \equiv \mu(\hat{f})$.

At outcome (μ^1, u^1), agents $\tilde{w}$ and f are single and $u^1_{\tilde{w}} = u^1_f = 0$. Furthermore, $u^1_w + u^1_f = u_w - 1 < v(w, f)$. This implies that pair (w, f) can block outcome (μ^1, u^1) with payoffs $u^2_w = u^1_w = u_w - 1$ and $u^2_f = \hat{u}_f = u_f + 1$. We obtain outcome (μ^2, u^2) from outcome (μ^1, u^1) by matching blocking pair (w, f) with payoffs $u^2_w \geq u^1_w$ and $u^2_f > u^1_f$. Outcome (μ^2, u^2) equals outcome $(\mu, \hat{u})$ with agents $\tilde{w}$ and $\hat{f}$ unmatched.

Inequality (3) together with our assumption that $u_w > u'_w$ implies $u'_w + u_{\hat{f}} < v(w, \hat{f})$. Outcome (μ', u') being stable implies $v(w, \hat{f}) \leq u'_w + u'_{\hat{f}}$. Hence, $u_{\hat{f}} < u'_{\hat{f}}$ and $u_{\hat{f}} + 1$ is a stable payoff for agent $\hat{f}$. Since (μ, u) is a stable outcome and $\tilde{w} = \mu(\hat{f})$, $u_{\tilde{w}} + u_{\hat{f}} = v(\tilde{w}, \hat{f})$ and $u_{\tilde{w}} - 1$ is a stable payoff for agent $\tilde{w}$. This implies that pair $(\tilde{w}, \hat{f})$ can block outcome (μ^2, u^2) with payoffs $u^3_{\tilde{w}} = u_{\tilde{w}} - 1$ and $u^3_{\hat{f}} = u_{\hat{f}} + 1$. We obtain outcome (μ, u^3) from outcome (μ^2, u^2) by matching blocking pair $(\tilde{w}, \hat{f})$ with payoffs $u^3_{\tilde{w}} \geq u^2_{\tilde{w}}$ and $u^3_{\hat{f}} > u^2_{\hat{f}}$.

Let $T^3 = \{w, \tilde{w}, f, \hat{f}\}$. Note that outcome (μ, u^3) is such that for all $i \notin T^3$, $u^3_i = u_i$ and

$$\begin{array}{lll} u_w > & u^3_w = u_w - 1 & \geq u'_w, \\ u_{\tilde{w}} > & u^3_{\tilde{w}} = u_{\tilde{w}} - 1 & \geq u'_{\tilde{w}}, \\ u_f < & u^3_f = u_f + 1 & \leq u'_f, \\ u_{\hat{f}} < & u^3_{\hat{f}} = u_{\hat{f}} + 1 & \leq u'_{\hat{f}}. \end{array}$$

Therefore, outcome (μ, u^3) is payoff closer to (μ', u') than (μ, u) is. Furthermore, since $u^3_w + u^3_f = u_w + u_f$ and $u^3_{\tilde{w}} + u^3_{\hat{f}} = u_{\tilde{w}} + u_{\hat{f}}$, the stability of (μ, u) implies that there is no blocking pair within the set T^3. By the definition of outcome (μ, u^3) and the stability of (μ, u) there is also no blocking pair within the set $(W \cup F) \setminus T^3$.

If (μ, u^3) is stable, then we are done. If not, then we construct a new outcome (μ, u^6) payoff closer to (μ', u') than (μ, u) is and a set $T^6 \supsetneqq T^3$. More generally, we assume that we have obtained an outcome (μ, u^k) and a set T^k with the following properties

(i) the set T^k contains an equal number of workers and firms such that

$$\begin{array}{llll} \text{for all } w \in T^k, & u_w > & u^k_w = u_w - 1 & \geq u'_w \text{ and} \\ \text{for all } f \in T^k, & u_f < & u^k_f = u_f + 1 & \leq u'_f, \end{array}$$

(ii) for all $i \notin T^k$, $u^k_i = u_i$,

(iii) outcome (μ, u^k) is payoff closer to (μ', u') than (μ, u) is, and

(iv) there exist no blocking pairs within the set T^k or the set $(W \cup F) \setminus T^k$.

Assume that (μ, u^k) is not stable. Note that outcome (μ, u^k) is individually rational. Hence, there exists a blocking pair $(\hat{w}, \hat{f}) \in W \times F$ with

$$u^k_{\hat{w}} + u^k_{\hat{f}} < v(\hat{w}, \hat{f}). \tag{5}$$

Assume $\hat{w} \notin T^k$ and recall that for $f \in T^k$, $u^k_f = u_f + 1 > u_f$ and for $i \in (W \cup F) \setminus T^k$, $\hat{u}_i = u_i$. Then, (5) implies $u_{\hat{w}} + u_{\hat{f}} < v(\hat{w}, \hat{f})$, contradicting the stability of outcome (μ, u). Hence, $\hat{w} \in T^k$ and $\hat{f} \notin T^k$. Let $\tilde{f} \equiv \mu(\hat{w})$.

Since (μ, u) is stable,

$$u_{\hat{w}} + u_{\hat{f}} \geq v(\hat{w}, \hat{f}). \tag{6}$$

Using $u^k_{\hat{w}} = u_{\hat{w}} - 1$ and $u^k_{\hat{f}} = u_{\hat{f}}$, inequality (5) reads as

$$u_{\hat{w}} + u_{\hat{f}} - 1 < v(\hat{w}, \hat{f}). \tag{7}$$

Hence, inequalities (6) and (7) together imply that

$$u_{\hat{w}} + u_{\hat{f}} = v(\hat{w}, \hat{f}). \tag{8}$$

Now inequalities (5) and (8) together imply that pair $(\hat{w}, \hat{f})$ can block outcome (μ, u^k) with payoffs $u^{k+1}_{\hat{w}} = u^k_{\hat{w}} = u_{\hat{w}} - 1$ and $u^{k+1}_{\hat{f}} = u^k_{\hat{f}} + 1 = u_{\hat{f}} + 1$. We obtain outcome (μ^{k+1}, u^{k+1}) from outcome (μ, u^k) by matching blocking pair $(\hat{w}, \hat{f})$ with payoffs $u^{k+1}_{\hat{w}} \geq u^k_{\hat{w}}$ and $u^{k+1}_{\hat{f}} > u^k_{\hat{f}}$.

Assume that $\mu(\hat{f}) = \hat{f}$. Then, by Lemma 1, $u_{\hat{f}} = u'_{\hat{f}} = 0$. Thus, by equation (8), $u_{\hat{w}} = v(\hat{w}, \hat{f})$. Then, our assumption (i) that $u_{\hat{w}} > u'_{\hat{w}}$ implies $v(\hat{w}, \hat{f}) = u_{\hat{w}} > u'_{\hat{w}} = u'_{\hat{w}} + u'_{\hat{f}}$, contradicting the stability of (μ, u'). Hence, $\mu(\hat{f}) \neq \hat{f}$ and $\mu(\hat{f}) \in W$. Let $\tilde{w} \equiv \mu(\hat{f})$.

At outcome (μ^{k+1}, u^{k+1}), agents $\tilde{w}$ and $\tilde{f}$ are single and $u^{k+1}_{\tilde{w}} = u^{k+1}_{\tilde{f}} = 0$. Furthermore, $u^{k+1}_{\hat{w}} + u^{k+1}_{\tilde{f}} = u_{\hat{w}} - 1 < v(\hat{w}, \tilde{f})$. This implies that pair $(\hat{w}, \tilde{f})$ can block outcome (μ^{k+1}, u^{k+1}) with payoffs $u^{k+2}_{\hat{w}} = u^{k+1}_{\hat{w}} = u_{\hat{w}} - 1$ and $u^{k+2}_{\tilde{f}} = u^k_{\tilde{f}} = u_{\tilde{f}} + 1$. We obtain outcome (μ^{k+2}, u^{k+2}) from outcome (μ^{k+1}, u^{k+1}) by matching blocking pair $(\hat{w}, \tilde{f})$ with payoffs $u^{k+2}_{\hat{w}} \geq u^{k+1}_{\hat{w}}$ and $u^{k+2}_{\tilde{f}} > u^{k+1}_{\tilde{f}}$. Outcome (μ^{k+2}, u^{k+2}) equals outcome (μ, u^k) with agents $\tilde{w}$ and $\hat{f}$ unmatched.

Inequality (7) together with our assumption that $u_{\hat{w}} > u'_{\hat{w}}$ implies $u'_{\hat{w}} + u_{\hat{f}} < v(\hat{w}, \hat{f})$. Outcome (μ', u') being stable implies $v(\hat{w}, \hat{f}) \leq u'_{\hat{w}} + u'_{\hat{f}}$. Hence, $u_{\hat{f}} < u'_{\hat{f}}$ and $u_{\hat{f}} + 1$ is a stable payoff for agent $\hat{f}$. Since (μ, u) is a stable outcome and $\tilde{w} = \mu(\hat{f})$, $u_{\tilde{w}} + u_{\hat{f}} = v(\tilde{w}, \hat{f})$ and $u_{\tilde{w}} - 1$ is a stable payoff for agent $\tilde{w}$. This implies that pair $(\tilde{w}, \hat{f})$ can block outcome (μ^{k+2}, u^{k+2}) with payoffs $u^{k+3}_{\tilde{w}} = u_{\tilde{w}} - 1$ and $u^{k+3}_{\hat{f}} = u_{\hat{f}} + 1$. We obtain outcome (μ, u^{k+3}) from outcome (μ^{k+2}, u^{k+2}) by matching blocking pair $(\tilde{w}, \hat{f})$ with payoffs $u^{k+3}_{\tilde{w}} \geq u^{k+2}_{\tilde{w}}$ and $u^{k+3}_{\hat{f}} > u^{k+2}_{\hat{f}}$.

Let $T^{k+3} = T^k \cup \{\tilde{w}, \hat{f}\}$. Note that outcome (μ, u^3) and T^{k+3} are such that

(i) the set T^{k+3} contains an equal number of workers and firms such that

$$\begin{array}{lll} \text{for all } w \in T^{k+3}, & u_w > & u_w^k = u_w - 1 \quad \geq u'_w \text{ and} \\ \text{for all } f \in T^{k+3}, & u_f < & u_f^k = u_f + 1 \quad \leq u'_f, \end{array}$$

(ii) for all $i \notin T^{k+3}$, $u_i^k = u_i$, and

(iii) outcome (μ, u^k) is payoff closer to (μ', u') than (μ, u) is.

Furthermore, since for any $w, f \in T^{k+3}$, $u_w^{k+3} + u_f^{k+3} = u_w + u_f$, the stability of (μ, u) implies that there is no blocking pair within the set T^{k+3}. By the definition of outcome (μ, u^{k+3}) and the stability of (μ, u) there is also no blocking pair within the set $(W \cup F) \setminus T^{k+3}$. This implies property (iv) of outcome (μ, u^3) and T^{k+3}.

Since $T^{k+3} \supsetneqq T^k$ and the set of agents is finite, the blocking path we are constructing must be finite. By construction, the resulting blocking path $(\mu, \hat{u}), (\mu^1, u^1), ..., (\mu, \bar{u})$ ends in a stable outcome $(\mu, \bar{u})$ that is payoff closer to (μ', u') than (μ, u) is. □

Proof of Theorem 3. Let (W, F, v) be an assignment problem and $(\mu, u), (\mu', u) \in \mathcal{S}(W, F, v)$ with $\mu \neq \mu'$ and u such that for all $i \neq \mu(i)$, $u_i + u_{\mu(i)} > 0$. Take a pair (w, f) such that $\mu(w) = f$ and $\mu'(w) \neq f$. By assumption $u_w + u_f > 0$. Without loss of generality, assume $u_w > 0$ (if $u_w = 0$, then $u_f > 0$ and we would use f instead of w). As $u_w > 0$, $\mu'(w) \neq w$. For any matching $\bar{\mu} \in \mathcal{M}(W, F)$, define $T(\bar{\mu}) \equiv \{i \in W \cup F \mid \bar{\mu}(i) = \mu'(i)\}$.

Obtain outcome $(\mu, \hat{u})$ from (μ, u) by (re)matching agents w, f such that $\hat{u}_w = u_w - 1$ and $\hat{u}_f = u_f + 1$. Hence, agents w and f stay matched and agent w makes a 1-error.

Recall that $\mu(w) \neq \mu'(w)$. Let $f_1 = \mu'(w)$. Since w and f_1 are optimal partners at μ', $u_w + u_{f_1} = v(w, f_1)$. Since $\hat{u}_w = u_w - 1$, pair (w, f_1) can block outcome $(\mu, \hat{u})$ with payoffs $u_w^1 = \hat{u}_w + 1 = u_w$ and $u_{f_1}^1 = \hat{u}_{f_1} = u_{f_1}$. We obtain outcome (μ^1, u^1) from outcome $(\mu, \hat{u})$ by matching blocking pair (w, f_1) with payoffs $u_w^1 > \hat{u}_w$ and $u_{f_1}^1 = \hat{u}_{f_1}$. Note that since the blocking pair is matched at μ', $T(\mu^1) \supsetneqq T(\mu)$. Hence, $m(\mu', \mu^1) > m(\mu', \mu)$ and outcome (μ^1, u^1) is match closer to outcome (μ', u) than outcome (μ, u) is. Agent f being single at μ^1 implies $u_f^1 = 0$.

If (μ^1, u^1) is stable, then f is not able to block with agent w and hence $u_f = 0$ (recall that $\mu(f) = w$, $u_w^1 = u_w$, and $u_w + u_f = v(w, f)$). Thus, $u_f^1 = u_f = 0$. If $\mu(f_1) \in W$, then agent $\mu(f_1)$ being single at μ^1 implies $u_{\mu(f_1)}^1 = 0$. Similarly as for agent f, agent $\mu(f_1)$ not being able to block with agent f_1 implies that $u_{\mu(f_1)}^1 = u_{\mu(f_1)} = 0$. This implies $u^1 = u$ and we have obtained a stable outcome (μ^1, u) that is match closer to (μ', u) than (μ, u) (we are done).

Assume that (μ^1, u^1) is not stable. Note that outcome (μ^1, u^1) is individually rational. Hence, there exists a blocking pair $(\hat{w}, \hat{f}) \in W \times F$ with

$$u_{\hat{w}}^1 + u_{\hat{f}}^1 < v(\hat{w}, \hat{f}). \tag{9}$$

Note that for all $i \neq f$ and $i \neq \mu(f_1)$ (if $\mu(f_1) \in W$), $u_i^1 = u_i$. Outcome (μ, u) being stable then implies that μ^1-single agent $f = \mu(w)$ or μ^1-single agent $\mu(f_1)$ (if $\mu(f_1) \in W$) participate in such a blocking pair. Furthermore, $u_f > 0$ or $u_{\mu(f_1)} > 0$ (if $\mu(f_1) \in W$).[8]

[8] If $u_f = 0$ and $u_{\mu(f_1)} = 0$ (if $\mu(f_1) \in W$), then $u^1 = u$ and (μ^1, u^1) is stable: a contradiction.

If $\mu'(f) = \mu(f_1)$, then outcome (μ, u) being stable and inequality (9) imply $u_{\mu(f_1)} + u_f = v(\mu(f_1), f) > 0$. Then, we obtain outcome (μ^2, u^2) from outcome (μ^1, u^1) by matching blocking pair $(\mu(f_1), f)$ with payoffs $u^2_{\mu(f_1)} = u_{\mu(f_1)} \geq u^1_{\mu(f_1)} = 0$ and $u^2_f = u_f \geq u^1_f$ with one blocking inequality being strict. Note that since the blocking pair is matched at μ', $T(\mu^2) \supsetneqq T(\mu^1) \supsetneqq T(\mu)$. Hence, $m(\mu', \mu^2) > m(\mu', \mu)$ and outcome (μ^2, u^2) is match closer to outcome (μ', u) than outcome (μ, u) is. Furthermore, $u^2 = u$ and outcome (μ^2, u) is stable (we are done).

If $\mu'(f) \neq \mu(f_1)$ and $u_f > u^1_f = 0$, then $f \neq \mu'(f)$ and $u_{\mu'(f)} + u_f = v(\mu'(f), f) > 0$. Then, we obtain outcome (μ^2, u^2) from outcome (μ^1, u^1) by matching blocking pair $(\mu'(f), f)$ with payoffs $u^2_{\mu'(f)} = u_{\mu'(f)} = u^1_{\mu'(f)}$ and $u^2_f = u_f > u^1_f = 0$. Note that since the blocking pair is matched at μ', $T(\mu^2) \supsetneqq T(\mu^1) \supsetneqq T(\mu)$. Hence, $m(\mu', \mu^2) > m(\mu', \mu)$ and outcome (μ^2, u^2) is match closer to outcome (μ', u) than outcome (μ, u) is.

If $\mu'(f) \neq \mu(f_1)$ and $u_f = u^1_f = 0$, then set $(\mu^2, u^2) \equiv (\mu^1, u^1)$. Since $T(\mu^2) = T(\mu^1) \supsetneqq T(\mu)$, (μ^2, u^2) is match closer to outcome (μ', u) than outcome (μ, u) is.

If $\mu'(f) \neq \mu(f_1)$, $\mu(f_1) \in W$, and $u_{\mu(f_1)} > u^2_{\mu(f_1)} = 0$, then $\mu(f_1) \neq \mu'(\mu(f_1))$ and $u_{\mu(f_1)} + u_{\mu'(\mu(f_1))} = v(\mu(f_1), \mu'(\mu(f_1))) > 0$. Then, we obtain outcome (μ^3, u^3) from outcome (μ^2, u^2) by matching blocking pair $(\mu(f_1), \mu'(\mu(f_1)))$ with payoffs $u^3_{\mu(f_1)} = u_{\mu(f_1)} > u^2_{\mu(f_1)} = 0$ and $u^3_{\mu'(\mu(f_1))} = u_{\mu'(\mu(f_1))} = u^2_{\mu'(\mu(f_1))}$. Note that since the blocking pair is matched at μ', $T(\mu^3) \supsetneqq T(\mu^2) \supsetneqq T(\mu)$. Hence, $m(\mu', \mu^3) > m(\mu', \mu)$ and outcome (μ^3, u^3) is match closer to outcome (μ', u) than outcome (μ, u) is.

If $\mu'(f) \neq \mu(f_1)$, $\mu(f_1) \in W$, and $u_{\mu(f_1)} = u^2_{\mu(f_1)} = 0$ or if $\mu(f_1) = f_1$, then set $(\mu^3, u^3) \equiv (\mu^2, u^2)$. Since $T(\mu^3) = T(\mu^2) \supsetneqq T(\mu)$, (μ^3, u^3) is match closer to outcome (μ', u) than outcome (μ, u) is.

We will now describe how outcome (μ^3, u^3) can look given the various blocking possibilities discussed before. We have the following ways in which matching μ^3 can have been obtained. First, matching μ^3 can be obtained from matching μ by agent w matching with agent f_1 and agent f matching with agent $\mu'(f) \in W$. Second, matching μ^3 can be obtained from matching μ by agent w matching with agent f_1 and by agent $\mu(f_1) \in W$ matching with agent $\mu'(\mu(f_1)) \in F$. Third, matching μ^3 can be obtained from matching μ by agent w matching with agent f_1, agent f matching with agent $\mu'(f) \in W$, and by agent $\mu(f_1) \in W$ matching with agent $\mu'(\mu(f_1)) \in F$.

Hence, by relabeling the names of the agents, we can describe the first two cases as follows: given matching μ such that $\mu(w_1) = f_1$ and $\mu(w_2) = f_2$ and matching μ' such that $\mu'(w_1) = f_2$ and $\mu'(w_2) = f_3$, we obtain matching μ^3 by unmatching the "μ-pairs" (w_1, f_1) and (w_2, f_2) and rematching them into "μ'-pairs" (w_1, f_2) and (w_2, f_3). Outcome (μ^3, u^3) is such that $u^3_{f_1} = 0$, $u^3_{\mu(f_3)} = 0$ if $\mu(f_3) \neq f_3$, and for all remaining agents i, $u^3_i = u_i$. The third case is similar but it contains an additional pair of agents: given matching μ such that $\mu(w_1) = f_1$, $\mu(w_2) = f_2$, and $\mu(w_3) = f_3$ and matching μ' such that $\mu'(w_1) = f_2$ and $\mu'(w_2) = f_3$, and $\mu'(w_3) = f_4$, we obtain matching μ^3 by unmatching the "μ-pairs" (w_1, f_1), (w_2, f_2), and (w_3, f_3) and rematching them into "μ'-pairs" (w_1, f_2), (w_2, f_3), and (w_3, f_4). Outcome (μ^3, u^3) is such that $u^3_{f_1} = 0$, $u^3_{\mu(f_4)} = 0$ if $\mu(f_4) \neq f_4$, and for all remaining agents i, $u^3_i = u_i$.

For the remainder of the proof, we assume that we have obtained an outcome $(\hat{\mu}^1, \hat{u}^1)$ as follows: for $k \geq 3$ and matching μ such that $\mu(w_1) = f_1$, $\mu(w_2) = f_2$, ..., $\mu(w_{k-1}) = f_{k-1}$ and matching μ' such that $\mu'(w_1) = f_2$, $\mu'(w_2) = f_3$, ..., $\mu'(w_{k-1}) = f_k$ we obtain matching $\hat{\mu}^1$ by unmatching the "μ-pairs" (w_1, f_1), (w_2, f_2), ..., (w_{k-1}, f_{k-1}) and rematching them into "μ'-pairs" (w_1, f_2), (w_2, f_3), ..., (w_{k-1}, f_k). Outcome $(\hat{\mu}^1, \hat{u}^1)$ is such that $\hat{u}^1_{f_1} = 0$, $\hat{u}^1_{\mu(f_k)} = 0$ if $\mu(f_k) \neq f_k$, and for all remaining agents i, $\hat{u}^1_i = u_i$. Our previously obtained outcome for (μ^3, u^3) corresponds to outcome $(\hat{\mu}^1, \hat{u}^1)$ with $k \in \{3, 4\}$. More generally, the following steps can be applied to outcome $(\hat{\mu}^1, \hat{u}^1)$ with any $k \geq 3$.

If $(\hat{\mu}^1, \hat{u}^1)$ is stable, then f_1 is not able to block with agent w_1 and hence $u_{f_1} = 0$ (recall that $\mu(f_1) = w_1$, $\hat{u}^1_{w_1} = u_{w_1}$, and $u_{w_1} + u_{f_1} = v(w_1, f_1)$). Thus, $\hat{u}^1_{f_1} = u_{f_1} = 0$. If $\mu(f_k) \in W$, then agent $\mu(f_k)$ being single at $\hat{\mu}^1$ implies $\hat{u}^1_{\mu(f_k)} = 0$. Similarly as for agent f_1, agent $\mu(f_k)$ not being able to block with agent f_k implies that $\hat{u}^1_{\mu(f_k)} = u_{\mu(f_k)} = 0$. This implies $\hat{u}^1 = u$ and we have obtained a stable outcome $(\hat{\mu}^1, u)$ that is match closer to (μ', u) than (μ, u) (we are done).

Assume that $(\hat{\mu}^1, \hat{u}^1)$ is not stable. Note that outcome $(\hat{\mu}^1, \hat{u}^1)$ is individually rational. Hence, there exists a blocking pair $(\hat{w}, \hat{f}) \in W \times F$ with

$$\hat{u}^1_{\hat{w}} + \hat{u}^1_{\hat{f}} < v(\hat{w}, \hat{f}). \tag{10}$$

Note that for all $i \neq f_1$ and $i \neq \mu(f_k)$ (if $\mu(f_k) \in W$), $\hat{u}^1_i = u_i$. Outcome (μ, u) being stable then implies that $\hat{\mu}^1$-single agent $f_1 = \mu(w_1)$ or $\hat{\mu}^1$-single agent $\mu(f_k)$ (if $\mu(f_k) \in W$) participate in such a blocking pair. Furthermore, $u_{f_1} > 0$ or $u_{\mu(f_k)} > 0$ (if $\mu(f_k) \in W$).[9]

If $\mu'(f_1) = \mu(f_k)$, then outcome (μ, u) being stable and inequality (10) imply $u_{\mu(f_k)} + u_{f_1} = v(\mu(f_k), f_1) > 0$. Then, we obtain outcome $(\hat{\mu}^2, \hat{u}^2)$ from outcome $(\hat{\mu}^1, \hat{u}^1)$ by matching blocking pair $(\mu(f_k), f_1)$ with payoffs $\hat{u}^2_{\mu(f_k)} = u_{\mu(f_k)} \geq \hat{u}^1_{\mu(f_k)} = 0$ and $\hat{u}^2_{f_1} = u_{f_1} \geq \hat{u}^1_{f_1}$ with one blocking inequality being strict. Note that since the blocking pair is matched at μ', $T(\hat{\mu}^2) \supsetneqq T(\hat{\mu}^1) \supsetneqq T(\mu)$. Hence, $m(\mu', \hat{\mu}^2) > m(\mu', \mu)$ and outcome $(\hat{\mu}^2, \hat{u}^2)$ is match closer to outcome (μ', u) than outcome (μ, u) is. Furthermore, $\hat{u}^2 = u$ and outcome $(\hat{\mu}^2, u)$ is stable (we are done).

If $\mu'(f_1) \neq \mu(f_k)$ and $u_{f_1} > \hat{u}^1_{f_1} = 0$, then $f_1 \neq \mu'(f_1)$ and $u_{\mu'(f_1)} + u_{f_1} = v(\mu'(f_1), f_1) > 0$. Then, we obtain outcome $(\hat{\mu}^2, \hat{u}^2)$ from outcome $(\hat{\mu}^1, \hat{u}^1)$ by matching blocking pair $(\mu'(f_1), f_1)$ with payoffs $\hat{u}^2_{\mu'(f_1)} = u_{\mu'(f_1)} = \hat{u}^1_{\mu'(f_1)}$ and $\hat{u}^2_{f_1} = u_{f_1} > \hat{u}^1_{f_1} = 0$. Note that since the blocking pair is matched at μ', $T(\hat{\mu}^2) \supsetneqq T(\hat{\mu}^1) \supsetneqq T(\mu)$. Hence, $m(\mu', \hat{\mu}^2) > m(\mu', \mu)$ and outcome $(\hat{\mu}^2, \hat{u}^2)$ is match closer to outcome (μ', u) than outcome (μ, u) is.

If $\mu'(f_1) \neq \mu(f_k)$ and $u_{f_1} = \hat{u}^1_{f_1} = 0$, then set $(\hat{\mu}^2, \hat{u}^2) \equiv (\hat{\mu}^1, \hat{u}^1)$. Since $T(\hat{\mu}^2) = T(\hat{\mu}^1) \supsetneqq T(\mu)$, $(\hat{\mu}^2, \hat{u}^2)$ is match closer to outcome (μ', u) than outcome (μ, u) is.

If $\mu'(f_1) \neq \mu(f_k)$, $\mu(f_k) \in W$, and $u_{\mu(f_k)} > \hat{u}^2_{\mu(f_k)} = 0$, then $\mu(f_k) \neq \mu'(\mu(f_k))$ and $u_{\mu(f_k)} + u_{\mu'(\mu(f_k))} = v(\mu(f_k), \mu'(\mu(f_k))) > 0$. Then, we obtain outcome $(\hat{\mu}^3, \hat{u}^3)$ from outcome $(\hat{\mu}^2, \hat{u}^2)$ by matching blocking pair $(\mu(f_k), \mu'(\mu(f_k)))$ with payoffs $\hat{u}^3_{\mu(f_k)} = u_{\mu(f_k)} > \hat{u}^2_{\mu(f_k)} = 0$ and $\hat{u}^3_{\mu'(\mu(f_k))} = u_{\mu'(\mu(f_k))} = \hat{u}^2_{\mu'(\mu(f_k))}$. Note that $T(\hat{\mu}^3) \supsetneqq T(\hat{\mu}^2) \supsetneqq T(\mu)$. Hence, $m(\mu', \hat{\mu}^3) > m(\mu', \mu)$ and outcome $(\hat{\mu}^3, \hat{u}^3)$ is match closer to outcome (μ', u) than outcome (μ, u) is.

[9] If $u_{f_1} = 0$ and $u_{\mu(f_k)} = 0$ (if $\mu(f_k) \in W$), then $\hat{u}^1 = u$ and $(\hat{\mu}^1, \hat{u}^1)$ is stable: a contradiction.

If $\mu'(f_1) \neq \mu(f_k)$, $\mu(f_k) \in W$, and $u_{\mu(f_k)} = \hat{u}^2_{\mu(f_k)} = 0$ or if $\mu(f_k) = f_k$, then set $(\hat{\mu}^3, \hat{u}^3) \equiv (\hat{\mu}^2, \hat{u}^2)$. Since $T(\hat{\mu}^3) = T(\hat{\mu}^2) \supsetneqq T(\mu)$, $(\hat{\mu}^3, \hat{u}^3)$ is match closer to outcome (μ', u) than outcome (μ, u) is.

Note that outcome $(\hat{\mu}^3, \hat{u}^3)$ has the same structure as outcome $(\hat{\mu}, \hat{u})$ but with a $\hat{k} > k$, i.e., for $\hat{k} > k$, matching μ is such that $\mu(w_1) = f_1$, $\mu(w_2) = f_2$, ..., $\mu(w_{\hat{k}-1}) = f_{\hat{k}-1}$ and matching μ' such that $\mu'(w_1) = f_2$, $\mu'(w_2) = f_3$, ..., $\mu'(w_{\hat{k}-1}) = f_{\hat{k}}$. We obtain matching $\hat{\mu}^3$ by unmatching the "μ-pairs" (w_1, f_1), (w_2, f_2), ..., $(w_{\hat{k}-1}, f_{\hat{k}-1})$ and rematching them into "μ'-pairs" (w_1, f_2), (w_2, f_3), ..., $(w_{\hat{k}-1}, f_{\hat{k}})$. Outcome $(\hat{\mu}^3, \hat{u}^3)$ is such that $\hat{u}^3_{f_1} = 0$, $\hat{u}^3_{\mu(f_{\hat{k}})} = 0$ if $\mu(f_{\hat{k}}) \neq f_{\hat{k}}$, and for all remaining agents i, $\hat{u}^3_i = u_i$.

Given that the set of agents is finite, repeating our arguments for each new and unstable outcome we end in stable outcome $(\hat{\mu}^l, \hat{u}^l)$ that is match closer to outcome (μ', u) than outcome (μ, u) is. □

Proof of Theorem 4. Let $\mathcal{S} := \mathcal{S}(W, F, v)$ and $\mathcal{O} := \mathcal{O}(W, F, v)$.

(i) Let $o \in \mathcal{S}$. Then, by definition of $\mathcal{S}$, no $(i, j) \in P(W, F)$ is a blocking pair for outcome o. Therefore $T(o, o) = 1$.

Note that (i) implies that for all $o \in \mathcal{O}$ and all $l \in \mathbb{N}$, $T^{l+1}(o, \mathcal{S}) \geq T^l(o, \mathcal{S})$.

(ii) By part (i), for $o \in \mathcal{S}$, $T(o, \mathcal{S}) = 1$. Let $o \in \mathcal{O} \setminus \mathcal{S}$. Then, by Theorem 1, there exists a blocking path of finite length from o to some $o' \in \mathcal{S}$. Let l_o be the length of the shortest such blocking path starting from o. We have that $T^{l_o}(o, \mathcal{S}) > 0$. Let $l = \max_{o \in \mathcal{O} \setminus \mathcal{S}} l_o$. Then, for all $o \in \mathcal{O}$, $T^l(o, \mathcal{S}) > 0$. Let $\xi = \min_{o \in \mathcal{O}} T^l(o, \mathcal{S})$. Then, for all $o \in \mathcal{O}$, $T^l(o, \mathcal{O} \setminus \mathcal{S}) < 1 - \xi$. Iterating, for all $o \in \mathcal{O}$ and all $n \in \mathbb{N}$, $T^{nl}(o, \mathcal{O} \setminus \mathcal{S}) < (1 - \xi)^n$. So $T^{nl}(o, \mathcal{O} \setminus \mathcal{S}) \to 0$ and therefore $T^{nl}(o, \mathcal{S}) \to 1$ as $n \to \infty$. □

Proof of Theorem 5. Let $\mathcal{S} := \mathcal{S}(W, F, v)$ and $\mathcal{SS} := \mathcal{SS}(W, F, v, c)$.

Assume $o = (\mu, u) \in \mathcal{SS}$, $\tilde{o} = (\tilde{\mu}, \tilde{u}) \in \mathcal{S}$. This implies that $(\mu, u) \in \mathcal{L}_{min}$, $(\tilde{\mu}, u) \in \mathcal{S}$. Take μ and unmatch all pairs $(i, \mu(i))$ satisfying $u_i + u_{\mu(i)} = 0$. Call this new matching μ'. Take $\tilde{\mu}$ and unmatch all pairs $(i, \tilde{\mu}(i))$ satisfying $u_i + u_{\tilde{\mu}(i)} = 0$. Call this new matching $\tilde{\mu}'$. Starting from $(\mu, u) = (\mu^1, u)$, obtain (μ^2, u) by unmatching some pair $(i, \mu(i))$ satisfying $u_i + u_{\mu(i)} = 0$. Note that $C((\mu^1, u), (\mu^2, u)) = 1$ and that $(\mu^2, u) \in \mathcal{S}$. Iterating, we eventually obtain $(\mu^{L_1}, u) = (\mu', u)$. From $(\mu^{L_1}, u) = (\mu', u)$, Theorem 3 shows that there exists $(\mu^{L_1+1}, u) \in \mathcal{S}$, such that $C((\mu^{L_1}, u), (\mu^{L_1+1}, u)) = 1$ and (μ^{L_1+1}, u) is match closer to $(\tilde{\mu}', u)$ than (μ^{L_1}, u) is. Iterating, we obtain a sequence $(\mu^{L_1}, u), (\mu^{L_1+1}, u), \ldots, (\mu^{L_2}, u) = (\tilde{\mu}', u)$. From $(\mu^{L_2}, u) = (\tilde{\mu}', u)$, obtain (μ^{L_2+1}, u) by matching some pair $(i, \mu(i))$ satisfying $u_i + u_{\tilde{\mu}(i)} = 0$. Note that $C((\mu^{L_2}, u), (\mu^{L_2+1}, u)) = 1$ and that $(\mu^{L_2+1}, u) \in \mathcal{S}$. Iterating, we eventually obtain $(\mu^{L_3}, u) = (\tilde{\mu}, u)$. From $(\mu, u^{L_3}) = (\tilde{\mu}, u)$, Theorem 2 shows that there exists $(\mu, u^{L_3+1}) \in \mathcal{S}$, such that $C((\mu, u^{L_3}), (\mu, u^{L_3+1})) = 1$ and (μ, u^{L_3+1}) is payoff closer to $(\tilde{\mu}, \tilde{u})$ than (μ, u^{L_3}) is. Iterating, we obtain a sequence $(\mu, u^{L_3}), (\mu, u^{L_3+1}), \ldots, (\mu, u^{L_4}) = (\tilde{\mu}, \tilde{u})$. So, we have obtained a sequence of stable states $(\mu, u) = (\mu^1, u^1), (\mu^2, u^2), \ldots, (\mu^{L_4}, u^{L_4}) = (\tilde{\mu}, \tilde{u})$ such that $C((\mu^l, u^l), (\mu^{l+1}, u^{l+1})) = 1$, $l = 1, \ldots, L_4 - 1$. Take g^* such that $\mathcal{V}(g^*) = \mathcal{V}_{min}((\mu, u))$. Remove edges exiting (μ^l, u^l), $l = 1, \ldots, L_4 - 1$, from g^*. This reduces $\mathcal{V}(.)$ by at least $L_4 - 1$. Add edges

$(\mu^l, u^l) \to (\mu^{l+1}, u^{l+1})$. This increases $\mathcal{V}(.)$ by at most $L_4 - 1$. Denote the new graph $\tilde{g}$ and note that $\tilde{g} \in \mathcal{G}((\tilde{\mu}, \tilde{u}))$. Therefore $\mathcal{V}_{min}((\tilde{\mu}, \tilde{u})) \leq \mathcal{V}_{min}((\mu, u))$. So it must be that $(\tilde{\mu}, \tilde{u}) \in \mathcal{L}_{min}$ and $(\tilde{\mu}, \tilde{u}) \in \mathcal{SS}$. □

Proof of Lemma 3. Let $u_i > 0$.

If $u_{\mu(i)} = 0$, then let (μ^1, u^1) be induced from (μ, u) by $\mu(i)$ becoming single. $u_i^1 = u_{\mu(i)}^1 = 0$. Then $c_{(\mu(i),\mu(i))}((\mu, u), (\mu^1, u^1)) = \delta$. Now, $(i, \mu(i))$ is a blocking pair for (μ^1, u^1). Let (μ^2, u^2) be induced from (μ^1, u^1) by $(i, \mu(i))$ matching with payoffs $u_i^2 = u_i - 1$ and $u_{\mu(i)}^2 = u_{\mu(i)} + 1$. (μ^2, u^2) is exactly the outcome that could be obtained from (μ, u) by a 1-error by i.

If $u_{\mu(i)} > 0$, then as $i \in \Delta_0$ implies $\mu(i) \in \Delta_0$, there exists an optimal matching $\mu^* \neq \mu$ such that $\mu^*(\mu(i)) \neq i$. Moreover, by Lemma 1 we know that $\mu^*(\mu(i)) \neq \mu(i)$. Let (μ^1, u^1) be induced from (μ, u) by $(\mu^*(\mu(i)), \mu(i))$ matching with payoffs $u_{\mu^*(\mu(i))}^1 = u_{\mu^*(\mu(i))}$ and $u_{\mu(i)}^1 = u_{\mu(i)}$. Now, $\mu^1(i) = i$, $u_i^1 = 0$, so $(i, \mu(i))$ is a blocking pair for (μ^1, u^1). Let (μ^2, u^2) be induced from (μ^1, u^1) by $(i, \mu(i))$ matching, $u_i^2 = u_i - 1$, $u_{\mu(i)}^2 = u_{\mu(i)} + 1$. If $\mu(\mu^*(\mu(i))) = \mu^*(\mu(i))$, then let $(\mu^3, u^3) = (\mu^2, u^2)$. Otherwise, $(\mu^*(\mu(i)), \mu(\mu^*(\mu(i))))$ is a blocking pair for (μ^2, u^2) (by the assumption that for all $j \neq \mu(j)$, $u_j + u_{\mu(j)} > 0$). Let (μ^3, u^3) be induced from (μ^2, u^2) by $(\mu^*(\mu(i)), \mu(\mu^*(\mu(i))))$ matching with payoffs $u_{\mu^*(\mu(i))}^3 = u_{\mu^*(\mu(i))}$ and $u_{\mu(\mu^*(\mu(i)))}^3 = u_{\mu(\mu^*(\mu(i)))}$. (μ^3, u^3) is exactly the outcome that could be obtained from (μ, u) by a 1-error by i. □

Proof of Theorem 6. Let $\mathcal{S} := \mathcal{S}(W, F, v)$ and $\mathcal{SS} := \mathcal{SS}(W, F, v, c)$.

Assume $o = (\mu, u) \in \mathcal{SS}$, $\tilde{o} = (\tilde{\mu}, \tilde{u}) \in \mathcal{S}$. This implies that $(\mu, u) \in \mathcal{L}_{min}$, $(\tilde{\mu}, u) \in \mathcal{S}$. Take μ and unmatch all pairs $(i, \mu(i))$ satisfying $u_i + u_{\mu(i)} = 0$. Call this new matching μ'. Take $\tilde{\mu}$ and unmatch all pairs $(i, \tilde{\mu}(i))$ satisfying $u_i + u_{\tilde{\mu}(i)} = 0$. Call this new matching $\tilde{\mu}'$. Starting from $(\mu, u) = (\mu^1, u)$, obtain (μ^2, u) by unmatching some pair $(i, \mu(i))$ satisfying $u_i + u_{\mu(i)} = 0$. Note that $C((\mu^1, u), (\mu^2, u)) = \delta$ and that $(\mu^2, u) \in \mathcal{S}$. Iterating, we eventually obtain $(\mu^{L_1}, u) = (\mu', u)$. Note that $\mu'(i) = j \neq i$ implies that $v(i, j) > 0$.

Similarly to the proof of Theorem 5, Theorem 3 shows the existence of a sequence of stable outcomes $(o^{L_1}, \ldots o^{L_2})$, $o^{L_1} = (\mu', u)$, $o^{L_2} = (\tilde{\mu}', u)$, such that $C(o^t, o^{t+1}) = \delta$, $t = L_1, \ldots, L_2 - 1$. This step is possible because the argument in Theorem 3 uses a 1-error at every step, which by Lemma 3 can be replicated by a 0-error.

From $o^{L_2} = (\tilde{\mu}', u)$, obtain o^{L_2+1} by matching some pair $(i, \tilde{\mu}(i))$ satisfying $u_i + u_{\tilde{\mu}(i)} = 0$. Note that $C((\mu^{L_2}, u), (\mu^{L_2+1}, u)) = \delta$ and that $(\mu^{L_2+1}, u) \in \mathcal{S}$. Iterating, we eventually obtain $o^{L_3} = (\tilde{\mu}, u)$.

The proof of Theorem 2 shows that if, for some $i \in \Delta_0$, $u_i > \tilde{u}_i$, then a 1-error by i can lead the process to a stable state which is closer to $(\tilde{\mu}, \tilde{u})$ than $(\tilde{\mu}, u)$ is. Lemma 3 shows that such a 1-error can be replicated by a 0-error. Therefore, there exists a sequence of stable outcomes $(o^{L_3}, \ldots o^{L_4})$, $o^{L_3} = (\tilde{\mu}, u)$, $o^{L_4} = (\tilde{\mu}, u^{L_4})$, $u_i^{L_4} = \tilde{u}_i$ for all $i \in \Delta_0$, $C(o^t, o^{t+1}) = \delta$, $t = L_3, \ldots, L_4 - 1$. An identical tree argument to the final part of the proof of Theorem 5 shows that $o^{L_4} \in \mathcal{SS}$ and completes the proof. □

The proofs of Theorems 7 and 8 follow from the discussion in the text and are omitted.

References

Agastya, M. (1997): "Adaptive Play in Multiplayer Bargaining Situations." *Review of Economic Studies*, 64(3): 411–26.

Agastya, M. (1999): "Perturbed Adaptive Dynamics in Coalition Form Games." *Journal of Economic Theory*, 89(2): 207–233.

Bergin, J. and Lipman, B. L. (1996): "Evolution with State-Dependent Mutations." *Econometrica*, 64(4): 943–56.

Biró, P., Bomhoff, M., Golovach, P. A., Kern, W., and Paulusma, D. (2013): "Solutions for the Stable Roommates Problem with Payments." In *Graph-Theoretic Concepts in Computer Science*, Lecture Notes in Computer Science, pages 69 –80. Springer Verlag.

Blume, L. E. (1993): "The Statistical Mechanics of Strategic Interaction." *Games and Economic Behavior*, 5(3): 387–424.

Chen, B., Fujishige, S., and Yang, Z. (2012): "Decentralized Market Processes to Stable Job Matchings with Competitive Salaries." Working paper, Department of Economics, University of York.

Crawford, V. P. and Knoer, E. M. (1981): "Job Matching with Heterogeneous Firms and Workers." *Econometrica*, 49(2): 437–450.

Demange, G. and Gale, D. (1985): "The Strategy Structure of Two-Sided Matching Markets." *Econometrica*, 53(4): 873–888.

Feldman, A. M. (1974): "Recontracting Stability." *Econometrica*, 42(1): 35–44.

Freidlin, M. I. and Wentzell, A. D. (1984): *Random Perturbations of Dynamical Systems*. Springer Verlag, first edition. (Second edition 1998.).

Green, J. R. (1974): "The Stability of Edgeworth's Recontracting Process." *Econometrica*, 42(1): 21–34.

Jackson, M. O. and Watts, A. (2002): "The Evolution of Social and Economic Networks." *Journal of Economic Theory*, 106(2): 265–295.

Klaus, B., Klijn, F., and Walzl, M. (2010): "Stochastic Stability for Roommate Markets." *Journal of Economic Theory*, 145(6): 2218 – 2240.

Klaus, B. and Payot, F. (2013): "Paths to Stability in the Assignment Problem." Cahier de recherches Économiques 13.14, University of Lausanne.

Maschler, M., Peleg, B., and Shapley, L. S. (1979): "Geometric Properties of the Kernel, Nucleolus, and Related Solution Concepts." *Mathematics of Operations Research*, 4(4): 303–338.

Nax, H. H. and Pradelski, B. S. R. (2013): "Decentralized Dynamics and Equitable Core Selection in Assignment Games." Working paper, University of Oxford.

Nax, H. H., Pradelski, B. S. R., and Young, H. P. (2013): "Decentralized Dynamics to Optimal and Stable States in the Assignment Game." In *Proceedings of the 52nd IEEE Conference on Decision and Control*, pages 2391–2397.

Newton, J. (2012a): "Coalitional Stochastic Stability." *Games and Economic Behavior*, 75(2): 842–54.

Newton, J. (2012b): "Recontracting and Stochastic Stability in Cooperative Games." *Journal of Economic Theory*, 147(1): 364–81.

Newton, J. and Sawa, R. (2013): "A One-Shot Deviation Principle for Stability in Matching Problems." Working Paper 2013-09, University of Sydney.

Núñez, M. and Rafels, C. (2008): "On the Dimension of the Core of the Assignment Game." *Games and Economic Behavior*, 64(1): 290–302.

Roth, A. E. and Sotomayor, M. A. O. (1990): *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press.

Sawa, R. (2013): "Coalitional Stochastic Stability in Games, Networks and Markets." Unpublished Manuscript, University of Aizu.

Sengupta, A. and Sengupta, K. (1996): "A Property of the Core." *Games and Economic Behavior*, 12(2): 266 – 273.

Serrano, R. and Volij, O. (2008): "Mistakes in Cooperation: the Stochastic Stability of Edgeworth's Recontracting." *Economic Journal*, 118(532): 1719–1741.

Shapley, L. S. and Shubik, M. (1971): "The Assignment Game I: The Core." *International Journal of Game Theory*, 1(1): 111–130.

van Damme, E. and Weibull, J. W. (2002): "Evolution in Games with Endogenous Mistake Probabilities." *Journal of Economic Theory*, 106(2): 296–315.

Young, H. P. (1993): "The Evolution of Conventions." *Econometrica*, 61(1): 57–84.

Young, H. P. (1998): *Individual Strategy and Social Structure*. Princeton University Press.