Journal of Mathematical Economics 62 (2016) 62–74

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

Stochastic stability in assignment problems☆

Bettina Klaus [a], Jonathan Newton [b,*]

[a] Faculty of Business and Economics (HEC), University of Lausanne, Internef 538, CH-1015 Lausanne, Switzerland

[b] School of Economics, University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 26 April 2015
Received in revised form
11 November 2015
Accepted 12 November 2015
Available online 23 November 2015

Keywords:
Assignment problem
(Core) stability
Decentralization
Stochastic stability

ABSTRACT

In a dynamic model of assignment problems, it is shown that 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.

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 et al., 2012; Biró et al., 2013; Klaus and Payot, forthcoming; Nax et al., 2013). 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.[1] 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).

In much the same way that perfectness concepts can be used to analyze the robustness of static equilibria to mistakes (Selten, 1975; Myerson, 1978), when perturbations are added to a dynamic process, more precise predictions about the long term behavior of the process can be obtained. Specifically, the invariant measure

☆ The authors would like to thank Heinrich Nax, Bary Pradelski, Jan Christoph Schlegel, Damian Sercombe, and Markus Walzl for detailed comments and questions. Bettina Klaus acknowledges financial support from the Swiss National Science Foundation (SNSF) and the University of Sydney (International Research Collaboration Award 2013). Jonathan Newton was supported by a Discovery Early Career Researcher Award funded by the Australian Research Council (Grant Number: DE130101768).

* Corresponding author.
E-mail addresses: bettina.klaus@unil.ch (B. Klaus), jonathan.newton@sydney.edu.au (J. Newton).

[1] There has long existed a literature on paths to stability in matching problems with non-transferable utility. See, for example, Roth and Vande Vate (1990), Diamantoudi et al. (2004), Klaus and Klijn (2007) and Kojima and Ünver (2008). It is therefore notable that similar problems have only recently begun to be addressed for matching problems with transferable utility.

http://dx.doi.org/10.1016/j.jmateco.2015.11.002
0304-4068/

of the perturbed process can place much greater weight on some equilibria of the unperturbed process than on others. As the probability of perturbations is taken to zero, the equilibria that have positive weight under the limiting invariant measure are known as *stochastically stable* equilibria. These are the outcomes that we would expect to observe most frequently in the long run when perturbations are rare. An example of such a process is the best-response dynamic (the dynamic justification for Nash equilibrium) together with some small probability that any given agent makes a mistake and does not play a best response. Recent experimental evidence supports perturbed best response as a behavioral rule, but is mixed as to whether mistakes depend strongly on associated payoff losses. Mäs and Nax (2014) find evidence that mistakes that cause greater payoff loss are less likely, whereas Lim and Neary (2013) find this not to be the case. The results of the current paper cover both of these possibilities.

A consequence of small deviations sufficing to move the process between core states is that stochastic stability is a weak selection concept. Results are first derived for two simple error processes, *uniform* and *stepped*, which occur in the literature. Following this, we define the class of *weakly payoff monotone* error processes. This is a large class of processes, under which errors that lead to greater payoff losses are (weakly) less likely. It is shown that for every process in this class, the results for either uniform or stepped errors pertain.

Uniform errors (see Young, 1993) are such that every error has the same (order of magnitude of) probability of occurring. Under uniform errors we show that 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 et al. (2010) which find no selection in marriage and roommate problems under uniform error processes.[2]

Payoff-dependent errors occur with probabilities that depend on the payoff loss incurred when they are made.[3] 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 3), 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 or to stepped errors (Theorem 8). If errors involving small payoff losses occur just as often as errors involving indifference, the result for uniform errors, Theorem 5, holds. If errors involving small payoff losses occur less often than errors involving indifference, then the results for stepped errors, Theorems 6 and 7, hold. Importantly, this latter class includes adaptations of popular choice rules, logit and probit choice, that are derived from random utility models.

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 modeled 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 (2015) have recently shown a maxmin selection result within the core for assignment games, a result discussed next.

2.2. Nax and Pradelski (2015)

Nax and Pradelski (2015) 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 (2015), Nax and Pradelski (2015) show that under this process the set of stochastically stable states is a subset of the least core (Maschler et al., 1979). If any agent has multiple optimal partners, then the least core equals the core, so 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 (2015) fails under the error processes 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, whereas the cited paper only considers errors with sufficient strength to cause a pair to break up. This restriction brings their model closer to models in the NTU literature, which is discussed next.

2.3. Selection in matching problems

Newton and Sawa (2015) 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

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

[3] The relation of such rules to uniform mistake models can be thought of as similar to the relation between the static concepts of Proper Equilibrium (Myerson, 1978) and Trembling Hand Perfect Equilibrium (Selten, 1975). In the former, mistakes associated with larger payoff losses are less likely, whereas in the latter there is no difference.

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.[4] We denote the set of pairs that agents in $W \cup F$ 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 $i \in W \cup F$.[5] 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 is *single at* μ. Thus, at any matching μ, the set of agents is partitioned into couples and singles.

A matching μ is *optimal* if, for all matchings μ', $\sum_{(i,j)\in\mu} v(i, j) \geq \sum_{(i,j)\in\mu'} v(i, j)$. 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 (μ, u) where μ is a matching and $u \in \mathbb{N}_0^{|W\cup F|}$ 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 $u_i = v(i, i) = 0$.[6]

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

Note the restriction of payoffs to integer values. This simplifies analysis, avoiding difficulties that can occur when analyzing either paths to stability or stochastic stability on uncountable state spaces (Klaus and Payot, forthcoming; Newton, 2015).

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* $v(w, f) - u_w - u_f > 0$. 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 exist, that is, for all $(w, f) \in W \times F$, we have $u_w + u_f \geq v(w, f)$. Let $\mathcal{S}(W, F, v)$ denote the *set of (core) stable outcomes*.

4. Blocking paths to stability

A *path* (for assignment problem (W, F, v)) is a sequence of outcomes $(\mu^1, u^1), \ldots, (\mu^k, u^k)$ such that for all $l \in \{1, \ldots, k-1\}$, 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.

A *blocking path* (for assignment problem (W, F, v)) is a path $(\mu^1, u^1), \ldots, (\mu^k, u^k)$ such that for all $l \in \{1, \ldots, k-1\}$, 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 $u_{w_l}^{l+1} \geq u_{w_l}^l$ and $u_{f_l}^{l+1} \geq u_{f_l}^l$ 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.

[4] 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 (forthcoming) and Nax et al. (2013).

[5] 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 justifiable by assuming that a smallest monetary unit of exchange exists.

[6] 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 $\sum_{i\in W\cup F} u_i = \sum_{(i,j)\in\mu} v(i, j)$.

Various recent papers (Chen et al., 2012; Biró et al., 2013; Klaus and Payot, forthcoming; 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 (μ, u) be an outcome. Then, there exists a blocking path $(\mu, u) = (\mu^1, u^1), \ldots, (\mu^k, u^k)$ that leads to stability, i.e., (μ^k, u^k) is stable.*

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 (μ, u), $(\hat{\mu}, \hat{u})$ is not a blocking path. 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 $\hat{u}_f = u_f - 1$.

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 $i \in W \cup F$, $| u'_i - \bar{u}_i | \leq | u'_i - u_i |$ 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 $(\mu, u), (\mu, \hat{u}), (\mu^1, u^1), \ldots, (\mu, \bar{u})$ 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) *$(\mu, \hat{u}), (\mu^1, u^1), \ldots, (\mu^k, u^k), (\mu, \bar{u})$ 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 the Appendix. 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 μ, μ', 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 similarity of 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 $(\mu, u), (\mu, \hat{u}), (\mu^1, u^1), \ldots, (\bar{\mu}, u)$ 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) *$(\mu, \hat{u}), (\mu^1, u^1), \ldots, (\bar{\mu}, u)$ 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 the Appendix. 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.

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

5. Stochastic stability

Similarly to the close relationship between Nash equilibrium and best response dynamics, the concept of pairwise stability in matching problems gives rise to a dynamic process in a very natural way. This has long been recognized in the literature on paths to stability in NTU matching problems (Roth and Vande Vate, 1990; Diamantoudi et al., 2004; Klaus and Klijn, 2007; Kojima and Ünver, 2008). The difference for TU problems such as the assignment game is that when a pair matches, they must choose how surplus is shared. To ensure consistency with the definition of pairwise stability under weak blocking, it must be that surplus is shared so that neither of the agents in the pair loses payoff and at least one of the agents gains payoff (Chen et al., 2012; Biró et al., 2013; Nax et al., 2013).

5.1. The unperturbed blocking dynamics and absorbing outcomes

For each assignment problem (W, F, v), we model the unperturbed blocking dynamics by a *Markov process* $(\mathcal{O}, T)$, where the *state space* is the set of outcomes $\mathcal{O}$ 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 $A(o, i, j) := \{o' \in \mathcal{O} \mid o'$ is obtained from o by matching $(i, j)\}$. Recall that possibly $i = j$ (in which case $u'_i = 0$).

Then, denote by $B(o, i, j) \subseteq A(o, i, j)$ 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. $B(o, i, j) := \{o' = (\mu', u') \in A(o, i, j) \mid u'_i \geq u_i,\ u'_j \geq u_j,$ 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 with probability mass function $g(\cdot)$ and full support on $P(W, F)$. Let o' be chosen randomly from a distribution with probability mass function $h_{A(o^t,i,j)}(\cdot)$ 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 by 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 (*Stability with Probability 1*). *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 $o \in \mathcal{O}$, $T^l(o, \mathcal{S}(W, F, v)) \to 1$ as $l \to \infty$.*

The proof is simple (see Appendix). Part (i) holds as by the definition of a stable outcome, there are no blocking pairs for any $o \in \mathcal{S}(W, F, v)$, 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 this argument implies that the probability of not being at a stable state approaches zero at an exponential rate. Further note that Theorem 4 holds independently of $g(\cdot)$ and $h_{A(\cdot,\cdot,\cdot)}(\cdot)$, as long as the full support assumptions are satisfied.

5.2. The perturbed blocking dynamics

When dealing with adaptive dynamics such as the one detailed above, it is common to consider perturbed variants of the process. The addition of perturbations can enable something to be said about the long term behavior of the process, regardless of initial conditions (Young, 1993). These perturbations can be considered as 'mistakes' made by agents. For example, two agents could form a couple and share surplus in such a way that one or both of them obtain a lower payoff than they did previously. The vast majority of the literature on perturbed adaptive dynamics in economics uses one of two types of error specification. Under uniform error specifications (Young, 1993), agents make mistakes and take payoff reducing actions with some uniform probability. Under payoff-dependent specifications, payoff reducing actions by an agent occur with a probability that is decreasing in his loss of payoff from taking the action in question. The most common form of payoff-dependence is the logit error specification (Blume, 1993), under which the probability of an error is log-linear in payoff loss. Other specifications such as probit (Myatt and Wallace, 2003; Dokumaci and Sandholm, 2011) are occasionally also considered. The logit and probit error specifications need not be considered to model mistakes, and can instead be considered as modeling the behavior of agents whose utility is subject to idiosyncratic payoff shocks (which follow extreme value and normal distributions respectively). 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 (2014) does both. Coalitional behavior is already an integral part of matching dynamics, as pairs constitute coalitions of size two.

Here, we start by examining two models of errors: a model with uniform errors and a model in which errors have a small amount of payoff dependence. It is then shown that results from these two cases extend to a large class of error models (weakly payoff monotone errors) that includes uniform and log-linear specifications.

Formally, 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' according to $h_{A(o^t,i,j)}(\cdot)$, 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 $\widetilde{T}^l(\cdot, \cdot)$, $l \in \mathbb{N}_+$, denote the transition probabilities associated with the perturbed blocking dynamics.

Possible errors include an agent accepting a lower payoff while remaining matched to the same partner, an agent leaving his current partner and matching with himself, an agent matching with a new partner and losing payoff, and an agent matching with a new partner without either of the parties gaining payoff. In short, an error is any transition by an agent or pair of agents that is not the result of a weak blocking. The results of this section shall only require errors that result in small or no loss in payoff for the erring agent or agents. However, this is not an assumption on the range of possible errors, but rather follows from Theorems 2 and 3, and the later Lemma 1.[7]

Note that for $\varepsilon > 0$, the perturbed blocking dynamics is irreducible and ergodic and therefore has a unique stationary distribution $\pi_\varepsilon(\cdot)$. By well known arguments (see Young, 1998), as $\varepsilon \to 0$, the limiting distribution $\pi_\varepsilon(\cdot) \to \pi(\cdot)$ 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 $\pi(\cdot)$ 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 cost functions $c_{(\cdot,\cdot)}(\cdot,\cdot)$. Therefore it is crucial that $c_{(\cdot,\cdot)}(\cdot,\cdot)$ is such that error probabilities are plausible. As noted above, our results cover a wide class of error specifications, including those most popular in the literature. To show this, we first give results for two specific specifications found in the literature, before showing that results for these two specifications extend to a broader class. To begin, we shall analyze the process with uniform errors.

Definition 1 (*Uniform Errors*)**.** An error process is uniform if, for all $o \in \mathcal{O}$, $(i, j) \in P(W, F)$, $o' \in A(o, i, j)$,

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

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.[8]

Theorem 5 (*No Selection With Uniform Errors*)**.** *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 two agents who are matched to one another in o separating to become singles in o'. The cost of separation will in general be different depending on which of the agents initiates the 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 $T \in \mathbb{N}_+$, $o^1 = o$, $o^T = o'$ 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 $\mathcal{S}(W, F, v)$ such that every vertex except for o has outdegree 1 and the graph has no cycles. For two outcomes $o', o'' \in \mathcal{S}(W, F, v)$, $o' \to o''$ 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

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

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 outcome in $\mathcal{S}(W, F, v)$ must be stochastically stable. The details of this proof are given in the Appendix.

The intuition behind the proof of Theorem 5 is that, from any stable outcome, any other stable outcome can be reached via a transition path on which other stable outcomes act as stepping stones. Moving from one stepping stone to the next only ever requires a single error (Theorems 2 and 3). In this sense, getting to any given stable outcome from anywhere else in the state space is never more difficult than getting to any other stable outcome. In the terminology of Nöldeke and Samuelson (1993), the set of all stable outcomes forms a *mutation connected component*. Thus, no stable outcome is any more stable to perturbations than any other stable outcome, and the set of stochastically stable outcomes equals the set of stable outcomes.

[7] This is in contrast to Nax and Pradelski (2015), who, when the process is at a stable outcome, in effect disallow errors that do not reach an outcome that is weakly blocked. That is, errors must be large enough to cause the break up of some existing partnership — errors which cause small shifts in surplus sharing between partners are not allowed.

[8] A corollary of the proof of Theorem 5 is that the mixing time for this perturbed process is $O(\varepsilon)$. The reason for this is that all of the important transitions between stable states are of order ε, and there do not exist more probable transitions away from any stable state.

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 agents, 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 (*Stepped Errors*)**.** An error process is stepped if, for all $o = (\mu, u), (i, j) \in P(W, F), o' = (\mu', u') \in A(o, i, j)$,

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

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 under the stepped error process.

Definition 3 (*0-errors*)**.** If $o = (\mu, u), o' = (\mu', u'), o' \neq o, (i, j) \in P(W, F), o' \in A(o, i, j), o' \notin B(o, i, j), u'_i = u_i, u'_j = u_j$, then we refer to a transition from o to o' as a 0-error.

Example 1 (*0-errors by Individuals*)**.** Let $W = \{w_1\}, F = \{f_1\}, v(w_1, f_1) = 10$. There is a unique optimal matching at which w_1, f_1 are matched, and an outcome is stable if and only if $u_{w_1} + u_{f_1} = 10$. From any stable outcome such that $u_{w_1}, u_{f_1} > 0$, all errors lead to some agent losing payoff, so no 0-errors are possible. However, from the stable outcome at which $u_{w_1} = 0, u_{f_1} = 10$, there exists a transition in which w_1 leaves f_1 to become single. Such a transition is not a weak blocking, but does not decrease the payoff of w_1. Such a transition is a 0-error. □

Example 2 (*0-errors by Pairs*)**.** Let $W = \{w_1, w_2\}, F = \{f_1\}, v(w_1, f_1) = v(w_2, f_1) = 10$. There are two optimal matchings at which f_1 is matched to w_1 and w_2 respectively, and all stable outcomes have $u_{w_1} = 0, u_{w_2} = 0, u_{f_1} = 10$. From the stable outcome at which w_1 and f_1 are matched, there exists a transition in which w_2 matches with f_1 to obtain payoffs of $u_{w_2} = 0, u_{f_1} = 10$. Such a transition is not a weak blocking, but does not decrease the payoff of w_2 or f_1. Such a transition is a 0-error. □

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.

Example 3 (*Selection with Stepped Errors: See Fig. 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 vectors of lowest and highest payoffs at stable outcomes are given by $\underline{u}_W = \underline{u}_F = (0, 0, 0), \bar{u}_W = \bar{u}_F = (20, 20, 10)$. 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 $i \in \{w_1, w_2\}, i \neq \mu(f_2)$. That is, f_2 and i match with one another but both receive the same payoff as before. Following this, 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 3 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 every 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 3 is that if 0-errors were costless, then core convergence would no longer apply in cases where there exist multiple optimal matchings.[9]

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 exist 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 1 (*Stepped Errors: Replication of 1-errors by 0-errors*)**.** *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 1 with the example in Fig. 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

[9] Although the existence of multiple optimal matchings is non-generic for continuous payoffs and in some sense unusual when payoffs are discrete, we do not emphasize this as it is easy to envisage plausible situations with multiple optimal matchings. For example, two sellers with identical outside options matching with a single buyer. The case of a unique optimal matching is dealt with in Theorem 7.

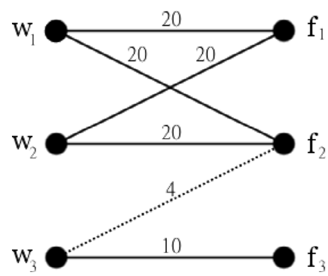

Fig. 1. A line between agents w_i, f_j indicates that $v(w_i, f_j) > 0$, with the value given above the line. Solid lines indicate pairings that can arise in some optimal matching.

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 1 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 (*Stepped Errors: No Selection for Agents with Multiple Optimal Partners*)**.** *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 optimal partners, 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. 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* $\mathcal{SS}(W, F, v, c) = \mathcal{S}(W, F, v)$*.*

In summary, any agent who is single at every optimal matching is single and obtains a payoff of zero at any stable outcome. Any agent who has different optimal partners can always have their partnership at a stable outcome broken by a 0-error, following which another stable outcome can be reached at which the agent's payoff or partner differs. Sequences of such transitions in which each step involves a single 0-error can move the process between any partner–payoff combinations that the agent has at any stable outcome. Therefore, if all agents are either always single or have multiple optimal partners, getting to any given stable outcome from anywhere else in the state space is never more difficult than getting to any other stable outcome, so all stable outcomes are stochastically stable.

We saw in Example 3 that if there exist agents who have a unique partner in all optimal matchings, then the influence of a latent outside option may lead to the set of stochastically stable outcomes being a strict subset of the set of stable outcomes. This strict selection in Example 3 was due to the latent outside option provided to w_3 by a potential matching with f_2 being strong enough to destabilize some of the stable payoffs of w_3, but not all of them. If the example is adjusted slightly so that the latent outside option can destabilize any stable payoff of w_3, then no-selection is again obtained.

Example 4 (*No Selection with Stepped Errors*)**.** In Example 3, we have $\mathcal{SS}(W, F, v, c) \subsetneq \mathcal{S}(W, F, v)$. However, if we alter the example by letting $v'(w_3, f_3) = 3$ and $v'(w_i, f_j) = v(w_i, f_j)$ for all $(w_i, f_j) \neq (w_3, f_3)$, then $\mathcal{SS}(W, F, v', c) = \mathcal{S}(W, F, v')$. This is due to the fact that w_3 and f_3 are still matched at any stable matching, but no matter how they divide the surplus of 3 between them, their partnership can still be broken up if f_2 becomes single, as f_2 can then weakly block with w_3 at payoffs $u_{w_3} = 4$, $u_{f_2} = 0$. As we saw in Example 3, it only takes a 0-error to cause f_2 to become single, so all possible payoff combinations can be traversed by means of 0-errors, and all stable outcomes are stochastically stable. □

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 $u'_i > u_i$. This implies that $\mu(i) \neq i$ [by Lemma 2(a) in the Appendix]. From (μ, u), following a 0-error by (i, j), let i rematch with $\mu(i)$ at payoffs $u_i + 1$, $u_{\mu(i)} - 1$ respectively [these payoffs must be possible as (μ, u') is an outcome and $u'_i > u_i$]. Then, if $\mu(j) \neq j$, let j rematch with $\mu(j)$ at payoffs u_j, $u_{\mu(j)}$ [it must be that $v(j, \mu(j)) > 0$ as if $v(j, \mu(j)) = 0$ then the matching obtained by unmatching j and $\mu(j)$ in μ would be optimal, contradicting there being a unique optimal matching]. The resultant outcome is the outcome that would have been obtained from (μ, u) had $\mu(i)$ made a 1-error. The resultant outcome 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 $Y \subset \mathcal{S}(W, F, v)$ 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.

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

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 5. Consider Example 3 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 so far 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 4 (*Weakly Payoff Monotone Errors*)**.** An error process is weakly payoff monotone if, for all $o = (\mu, u), (i, j) \in P(W, F), o' = (\mu', u') \in A(o, i, j)$,

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

where $g_1 : \mathbb{N}_0 \to \mathbb{R}_{++}, g_2 : \mathbb{N}_0^2 \to \mathbb{R}_{++}$ 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 made 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 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 (*Results for Weakly Payoff Monotone Errors*)**.** *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.*

Intuitively, as errors which lose a lot of payoff are (weakly) rarer than errors which lose little payoff, the most common transition paths between stable states will not include errors that cause large payoff losses to the agent or agents making them. Consider a move from one stable outcome to another stable outcome, possibly across stepping stones of intermediate stable outcomes. There exists such a sequence of transitions such that each step does not require more than a 1-error; however, sometimes a 0-error suffices. Therefore, results hinge on whether or not 1-errors are more costly than 0-errors.

Example 6 (*Logit Errors*)**.** 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 (2014) 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.[10] 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.[11] An error process is logit if, for all $o = (\mu, u), (i, j) \in P(W, F), o' = (\mu', u') \in A(o, i, j)$,

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

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

Example 7. Consider an error process whereby, for all $o = (\mu, u), (i, j) \in P(W, F), o' = (\mu', u') \in A(o, i, j)$,

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

Then, as thefactorial function is increasing and $0! = 1! = 1$, this rule falls into category (i) of Theorem 8, and therefore Theorem 5 holds. □

It is worth emphasizing that Theorem 8 is independent of the exact functional form of $g_1(\cdot)$ and $g_2(\cdot)$ and that such independence is unusual in pairwise dynamics. Recall that it is possible that an error by a pair $(i, j), i \neq j$, involves both i and j incurring a payoff loss. Consider two possible errors, an error by (w_1, f_1) such that w_1 loses payoff of 5 and f_1 loses payoff of 0, and an error by (w_2, f_2) such that w_2 loses a payoff of 3 and f_2 loses a payoff of 3. Which of these errors will have the lower cost depends on the functional forms of $g_1(\cdot)$ and $g_2(\cdot)$. The independence of Theorem 8 from these functional forms is due to the fact that the least cost transitions on which our results depend *never* involve errors in which more than one of the protagonists loses payoff.[12]

10 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 $\varepsilon^{l_1+l_2}$.

11 Specifically, starting from $o = (\mu, u)$, a mistake occurs with probability ε^{δ}, following which, setting $\varepsilon = e^{-\beta}$, alternative outcome $o' = (\mu', u')$ is accepted independently by players $k \in \{i, j\}$ with probabilities $\frac{e^{\beta u_k'}}{(e^{\beta u_k'} + e^{\beta u_k})}$. Taking limits as $\varepsilon \to 0$ and the product of the independent probabilities that a mistake will occur and an alternative accepted by both i and j, we obtain $c_{(i,j)}(o, o') = \delta + \sum_{k \in \{i,j\}} \max\{u_k - u_k', 0\}$. For more on coalitional choice under logit and probit rules, see Sawa (2014) and Newton and Sawa (2015).

12 This is in contrast to NTU matching problems. In such problems least cost transitions *can* involve errors in which two protagonists both lose payoff. Therefore, even for weakly payoff monotone error processes, results in NTU problems do depend on the exact functional form of $g_1(\cdot)$ and $g_2(\cdot)$. See Section 4 of Newton and Sawa (2015) for an example contrasting results for logit and probit choice rules.

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 stochastic stability results for a variety of perturbed blocking dynamics. Processes with uniform errors (Theorem 5) and with stepped errors (Theorems 6 and 7) are analyzed, and a large class of perturbed processes is 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.[13]

Appendix

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 2 (*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*).*

(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).*

Proof of Theorem 2. By statement of the theorem, (W, F, v) is an assignment problem and $(\mu, u), (\mu', u') \in \mathcal{S}(W, F, v)$ with $u \neq u'$ ((μ, u) is the starting stable outcome and (μ', u') is the target stable outcome). Hence, there exists $i \in W \cup F$ with $u_i \neq u'_i$. If for some $w \in W$ we have $u_w > u'_w$ (conversely, $u_w < u'_w$), then by Lemma 2(a) we have $\mu(w) = f \in F$. By Lemma 2(b), μ is an optimal matching, so by Lemma 2(c), $(\mu, u') \in \mathcal{S}(W, F, v)$. Thus, $u_w + u_f = u'_w + u'_f = v(w, f)$. Hence, $u_f < u'_f$ (conversely, $u_f > u'_f$). Without loss of generality we shall assume that there exists a $w \in W$ with $u_w > u'_w$, and therefore $\mu(w) = f \in F$ and $u_f < u'_f$ (in the alternative case we could simply repeat the following argument, switching the sets W and F).

Obtain outcome (μ, u^1) from (μ, u) by adjusting the payoffs of agents w, f such that $u^1_w = u_w - 1$ and $u^1_f = u_f + 1$. Hence, agents w and f stay matched, agent w makes a 1-error, and outcome (μ, u^1) is payoff closer to (μ', u') than (μ, u) is.

Let $R^1 := \{w, f\}$. Starting from R^1, (μ, u^1), we construct a blocking path that leads to successive (μ, u^k), R^k that satisfy the following properties.

(P-i) The set R^k is nonempty and contains an equal number of workers and firms such that

for all $w \in R^k$, $\quad u_w > u^k_w = u_w - 1 \geq u'_w$,

for all $f \in R^k$, $\quad u_f < u^k_f = u_f + 1 \leq u'_f$, and

for all $w \in R^k$, $\quad w \neq \mu(w) \in R^k$.

(P-ii) For all $i \notin R^k$, $u^k_i = u_i$.

Note that (P-i), (P-ii) and, for (P-iv), stability of (μ, u) further imply

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

(P-iv) there exist no blocking pairs for (μ, u^k) within the set R^k or the set $(W \cup F) \setminus R^k$.[14]

It is easy to check that (P-i, ii) are satisfied for R^1, (μ, u^1). Assume, as an inductive hypothesis, that a blocking path exists from (μ, u^1) to some (μ, u^k), R^k satisfying (P-i, ii). If (μ, u^k) is stable, we are done. If (μ, u^k) is not stable, there exists a blocking pair $(\hat{w}, \hat{f}) \in W \times F$ with

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

Case 1: $\hat{w} \notin R^k$. By (P-iv), $\hat{f} \in R^k$, therefore by (P-i, ii) we have $u^k_{\hat{w}} + u^k_{\hat{f}} = u_{\hat{w}} + u_{\hat{f}} + 1 > v(\hat{w}, \hat{f})$, with the final inequality following from stability of (μ, u). This contradicts (1).

Case 2: $\hat{w} \in R^k$. By (P-iv), $\hat{f} \notin R^k$. $\hat{w} \in R^k$ implies $u^k_{\hat{w}} = u_{\hat{w}} - 1$ and $\hat{f} \notin R^k$ implies $u^k_{\hat{f}} = u_{\hat{f}}$. Substituting into (1) we obtain $u_{\hat{w}} + u_{\hat{f}} - 1 < v(\hat{w}, \hat{f})$. However, stability of (μ, u) implies $u_{\hat{w}} + u_{\hat{f}} \geq v(\hat{w}, \hat{f})$, so it must be that

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

Together, (1) and (2) imply that we can obtain an outcome (μ^{k+1}, u^{k+1}) from outcome (μ, u^k) by matching blocking pair $(\hat{w}, \hat{f})$ with payoffs $u^{k+1}_{\hat{w}} = u^k_{\hat{w}} = u_{\hat{w}} - 1$ and $u^{k+1}_{\hat{f}} = u^k_{\hat{f}} + 1 = u_{\hat{f}} + 1$.

Case 2a: $\mu(\hat{f}) = \hat{f}$. By Lemma 2(a), $u_{\hat{f}} = u'_{\hat{f}} = 0$. Thus, by (2), $u_{\hat{w}} = v(\hat{w}, \hat{f})$. (P-i) and $\hat{w} \in R^k$ imply $u_{\hat{w}} > u'_{\hat{w}}$. So we have $v(\hat{w}, \hat{f}) = u_{\hat{w}} > u'_{\hat{w}} = u'_{\hat{w}} + u'_{\hat{f}}$, contradicting stability of (μ', u').

Case 2b: $\mu(\hat{f}) \in W$. This proof step is illustrated in Fig. 2. By (P-i), $\hat{w} \neq \mu(\hat{w}) \in R^k$. Let $\tilde{w} := \mu(\hat{f})$ and $\tilde{f} := \mu(\hat{w})$.

At outcome (μ^{k+1}, u^{k+1}), agents $\tilde{w}$ and $\tilde{f}$ are single and $u^{k+1}_{\tilde{w}} = u^{k+1}_{\tilde{f}} = 0$. As $\tilde{f} = \mu(\hat{w})$, we have $u_{\hat{w}} + u_{\tilde{f}} = v(\hat{w}, \tilde{f})$. Therefore $u^{k+1}_{\hat{w}} + u^{k+1}_{\tilde{f}} = u_{\hat{w}} - 1 < v(\hat{w}, \tilde{f})$ and $(\hat{w}, \tilde{f})$ is a blocking pair for (μ^{k+1}, u^{k+1}). We can obtain an 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 $u^{k+2}_{\hat{w}} = u^{k+1}_{\hat{w}} = u_{\hat{w}} - 1$ and $u^{k+2}_{\tilde{f}} = u_{\tilde{f}} + 1$. Note that outcome (μ^{k+2}, u^{k+2}) differs from (μ, u^k) only in that agents $\tilde{w}$ and $\hat{f}$ are now unmatched.

(P-i) and $\hat{w} \in R^k$ imply that $u'_{\hat{w}} < u_{\hat{w}}$. Together with (2) this 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, $u_{\hat{f}} < u'_{\hat{f}}$. Since (μ', u') is a stable outcome and, by Lemma 2(b), μ is an optimal matching, we have by Lemma 2(c) that (μ, u') is stable. As $\tilde{w} = \mu(\hat{f})$, it follows that

[13] Kojima and Ünver (2008) give a path to stability result for many-to-many matchings when one side of the market has responsive preferences and the other has substitutable preferences. Newton and Sawa (2015) give a stochastic stability result for many-to-one matchings with responsive preferences and show that this does not extend to substitutable preferences.

[14] If $(\hat{w}, \hat{f})$ is a blocking pair for (μ, u^k) in R^k (or in $(W \cup F) \setminus R^k$), then $u^k_{\hat{w}} + u^k_{\hat{f}} = u_{\hat{w}} + u_{\hat{f}} < v(u_{\hat{w}}, u_{\hat{f}})$ and $(\hat{w}, \hat{f})$ is a blocking pair for (μ, u); a contradiction.

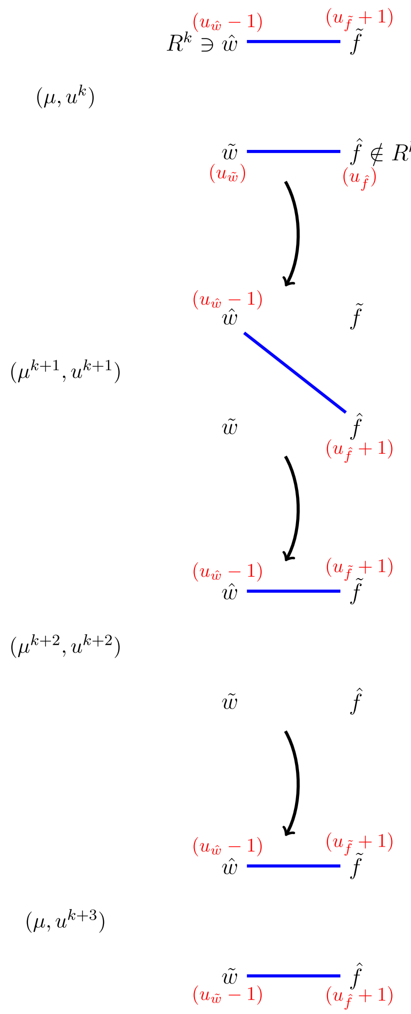

Fig. 2. Illustration of the inductive step in the proof of Theorem 2, Case 2b.

$u'_{\tilde{w}} + u'_{\hat{f}} = v(\tilde{w}, \hat{f})$. Together with $u_{\hat{f}} < u'_{\hat{f}}$, this implies $u'_{\tilde{w}} < u_{\tilde{w}}$, so it must be that $u_{\tilde{w}} > 0$. Therefore, since $\tilde{w}$ and $\hat{f}$ are single at μ^{k+2}, $0 = u^{k+2}_{\tilde{w}} + u^{k+2}_{\hat{f}} < u_{\tilde{w}} + u_{\hat{f}} = v(\tilde{w}, \hat{f})$ and $(\tilde{w}, \hat{f})$ is a blocking pair for (μ^{k+2}, u^{k+2}). We can obtain an outcome (μ, u^{k+3}) from outcome (μ^{k+2}, u^{k+2}) by matching blocking pair $(\tilde{w}, \hat{f})$ with payoffs $u^{k+3}_{\tilde{w}} = u_{\tilde{w}} - 1$ and $u^{k+3}_{\hat{f}} = u_{\hat{f}} + 1$.

Let $R^{k+3} = R^k \cup \{\tilde{w}, \hat{f}\}$. Note that outcome (μ, u^{k+3}) and R^{k+3} satisfy (P-i, ii).

Since from any unstable (μ, u^k) satisfying (P-i, ii) we can reach (μ, u^{k+3}), $R^{k+3} \supsetneqq R^k$ satisfying (P-i, ii), and the set of agents is finite, by iterating the above argument we must eventually attain a stable outcome. By (P-iii), this stable outcome is payoff closer to (μ', u') than (μ, u) is. □

Proof of Theorem 3. By statement of the theorem, (W, F, v) is an assignment problem and $(\mu, u), (\mu', u) \in \mathcal{S}(W, F, v)$ with $\mu \neq \mu'$ and u such that for all $i \neq \mu(i)$, $u_i + u_{\mu(i)} > 0$ ((μ, u) is the starting stable outcome and (μ', u) is the target stable outcome). 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).

Obtain outcome $(\mu, \hat{u})$ from (μ, u) by adjusting the payoffs of 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.

As $u_w > 0$, by Lemma 2(a) we have $\mu'(w) \neq w$. Let $f_1 = \mu'(w)$. Since (μ', u) is an outcome, $u_w + u_{f_1} = v(w, f_1)$. Hence, $\hat{u}_w + \hat{u}_{f_1} = u_w + u_{f_1} - 1 < v(w, f_1)$ and (w, f_1) is a blocking pair for $(\mu, \hat{u})$. We can obtain an outcome (μ^1, u^1) from outcome $(\mu, \hat{u})$ by matching blocking pair (w, f_1) with payoffs $u^1_w = \hat{u}_w + 1 = u_w$ and $u^1_{f_1} = \hat{u}_{f_1} = u_{f_1}$. Note that since $f_1 = \mu'(w)$, we have that $m(\mu', \mu^1) > m(\mu', \mu)$ [i.e. outcome (μ^1, u^1) is match closer to outcome (μ', u) than outcome (μ, u) is].

Starting from (μ^1, u^1) we construct a blocking path that leads to successive (μ^k, u^k) that satisfy the following properties.

(Q-i) There exist $w_1, \ldots, w_l \in W, f_1, \ldots, f_l \in F, l \geq 1$, such that

$$\mu^k(f_j) = \mu'(f_j) = w_j \quad \text{for } j = 1, \ldots, l,$$
$$\mu(f_j) = w_{j+1} \quad \text{for } j = 1, \ldots, l-1, \quad \text{and}$$

[Note that for (μ^1, u^1), we let $w_1 = w$]

(Q-ii) If $\mu(w_1) \in F$, then $\mu^k(\mu(w_1)) = \mu(w_1)$ and if $\mu(f_l) \in W$, then $\mu^k(\mu(f_l)) = \mu(f_l)$.
[i.e. any partners of w_1 or f_l at μ, are single at μ^k]

(Q-iii) For $i \in W \cup F$, unless $i = \mu(w_1) \in F$ or $i = \mu(f_l) \in W$, we have $u^k_i = u_i$. If $i = \mu(w_1) \in F$ or $i = \mu(f_l) \in W$, then $u^k_i = 0$.

(Q-iv) $m(\mu', \mu^k) > m(\mu', \mu)$.

Letting $w_1 = w$, it is easy to check that (Q-i, ii, iii, iv) are satisfied for (μ^1, u^1). Assume, as an inductive hypothesis, that a blocking path exists from (μ^1, u^1) to some (μ^k, u^k) satisfying (Q-i, ii, iii, iv). If (μ^k, u^k) is stable, we are done. If (μ^k, u^k) is not stable, then it must be that $\mu(w_1) \in F$, $u_{\mu(w_1)} > 0$ or $\mu(f_l) \in W$, $u_{\mu(f_l)} > 0$, as if neither of these held, then by (Q-iii), $u^k = u$ and (μ^k, u) would be stable.

Assume, without loss of generality, that $\mu(f_l) \in W$, $u_{\mu(f_l)} > 0$ (the alternative case is similar). By Lemma 2(a), this implies that $\mu(f_l) \neq \mu'(\mu(f_l)) \in F$. Since (μ', u) is an outcome, $u_{\mu(f_l)} + u_{\mu'(\mu(f_l))} = v(\mu(f_l), \mu'(\mu(f_l)))$.

Case 1: $\mu'(\mu(f_l)) = \mu(w_1)$. Since (μ', u) is an outcome, $u_{\mu(f_l)} > 0$ implies $u_{\mu(f_l)} + u_{\mu(w_1)} = v(\mu(f_l), \mu(w_1)) > 0$. As $u^k_{\mu(f_l)} = u^k_{\mu(w_1)} = 0$ by (Q-ii), $u^k_{\mu(f_l)} + u^k_{\mu(w_1)} < v(\mu(f_l), \mu(w_1))$ and $(\mu(f_l), \mu(w_1))$ is a blocking pair for (μ^k, u^k). We can obtain an outcome (μ^{k+1}, u^{k+1}) from outcome (μ^k, u^k) by matching blocking pair $(\mu(f_l), \mu(w_1))$ with payoffs $u^{k+1}_{\mu(f_l)} = u_{\mu(f_l)}$ and $u^{k+1}_{\mu(w_1)} = u_{\mu(w_1)}$. By (Q-ii) we then have $u^{k+1} = u$ and (μ^{k+1}, u^{k+1}) is stable. By $\mu'(\mu(f_l)) = \mu(w_1)$ and (Q-iv), we have $m(\mu', \mu^{k+1}) > m(\mu', \mu^k) > m(\mu', \mu)$, so we are done.

Case 2: $\mu'(\mu(f_l)) \neq \mu(w_1)$. This proof step is illustrated in Fig. 3. Since (μ', u) is an outcome, $u_{\mu(f_l)} + u_{\mu'(\mu(f_l))} = v(\mu(f_l), \mu'(\mu(f_l)))$. As $u_{\mu(f_l)} > 0$ by assumption, $u^k_{\mu(f_l)} = 0$ by (Q-ii), and $u^k_{\mu'(\mu(f_l))} = u_{\mu'(\mu(f_l))}$ by (Q-iii), we have $u^k_{\mu(f_l)} + u^k_{\mu'(\mu(f_l))} < u_{\mu(f_l)} + u_{\mu'(\mu(f_l))} = v(\mu(f_l), \mu'(\mu(f_l)))$ and $(\mu(f_l), \mu'(\mu(f_l)))$ is a blocking pair for (μ^k, u^k). We can obtain an outcome (μ^{k+1}, u^{k+1}) from outcome (μ^k, u^k) by matching blocking pair $(\mu(f_l), \mu'(\mu(f_l)))$ with payoffs $u^{k+1}_{\mu(f_l)} = u_{\mu(f_l)}$ and $u^{k+1}_{\mu'(\mu(f_l))} = u_{\mu'(\mu(f_l))}$. Denoting $\mu(f_l) =: w_{l+1}$ and $\mu'(\mu(f_l)) =: f_{l+1}$, we see that (μ^{k+1}, u^{k+1}) satisfies (Q-i, ii, iii, iv).

Given that the set of agents is finite, iterating we eventually reach some stable outcome (μ^n, u^n) satisfying $m(\mu', \mu^n) > m(\mu', \mu)$ and $u^n = u$. Note that the algorithm must terminate at a stable outcome, as eventually $(\mu^n, u^n) = (\mu', u)$ would have to be reached. □

Proof of Theorem 4. Let $\mathcal{S} := \mathcal{S}(W, F, v)$ and recall that $\mathcal{O}$ denotes the set of outcomes.

(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 $l \in \mathbb{N}$, $T^{l+1}(o, \mathcal{S}) \geq T^l(o, \mathcal{S})$.

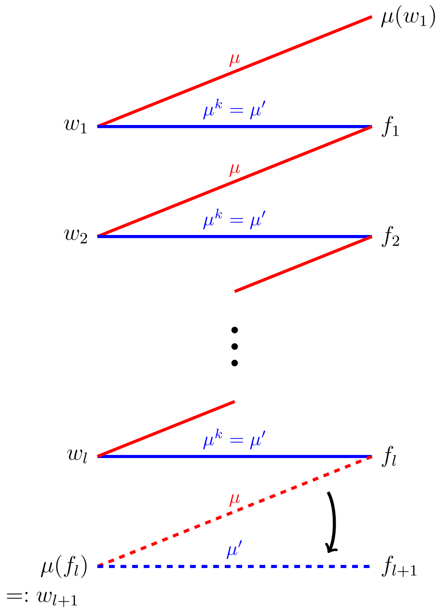

Fig. 3. Illustration of the inductive step in the proof of Theorem 3, Case 2.

(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 $\tilde{\mu}'$. By Lemma 2(b), matchings μ, μ' are optimal and so are $\tilde{\mu}, \tilde{\mu}'$.

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 by Lemma 2(c), $(\mu^2, u) \in \mathcal{S}$. Iterating, we eventually obtain $(\mu^{L_1}, u) = (\mu', u) \in \mathcal{S}$.

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 $(\mu^{L_1}, u), (\mu^{L_1+1}, u), \ldots, (\mu^{L_2}, u) = (\tilde{\mu}', u) \in \mathcal{S}$.

From $(\mu^{L_2}, u) = (\tilde{\mu}', u)$, obtain (μ^{L_2+1}, u) by matching some pair $(i, \tilde{\mu}(i))$ satisfying $u_i + u_{\tilde{\mu}(i)} = 0$. Note that $C((\mu^{L_2}, u), (\mu^{L_2+1}, u)) = 1$ and by Lemma 2(c), $(\mu^{L_2+1}, u) \in \mathcal{S}$. Iterating, we eventually obtain $(\mu^{L_3}, u) = (\tilde{\mu}, u) \in \mathcal{S}$.

From $(\mu^{L_3}, u) = (\tilde{\mu}, u)$, Theorem 2 shows that there exists $(\tilde{\mu}, u^{L_3+1}) \in \mathcal{S}$, such that $C((\tilde{\mu}, u^{L_3}), (\tilde{\mu}, u^{L_3+1})) = 1$ and $(\tilde{\mu}, u^{L_3+1})$ is payoff closer to $(\tilde{\mu}, \tilde{u})$ than $(\tilde{\mu}, u^{L_3})$ is. Iterating, we obtain a sequence $(\tilde{\mu}, u^{L_3}), (\tilde{\mu}, u^{L_3+1}), \ldots, (\tilde{\mu}, u^{L_4}) = (\tilde{\mu}, \tilde{u}) \in \mathcal{S}$.

Writing $(\mu^l, u^l) := (\mu^l, u)$ for $l = 1, \ldots, L_3$, and $(\mu^l, u^l) := (\tilde{\mu}, u^l)$ for $l = L_3 + 1, \ldots, L_4$, 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^* rooted at (μ, u) such that $\mathcal{V}(g^*) = \mathcal{V}_{min}((\mu, u))$. Remove edges exiting $(\mu^l, u^l), l = 1, \ldots, L_4 - 1$, from g^*. This reduces $\mathcal{V}(\cdot)$ by at least $L_4 - 1$. Add edges $(\mu^l, u^l) \to (\mu^{l+1}, u^{l+1})$. This increases $\mathcal{V}(\cdot)$ 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 $\mathcal{V}_{min}((\tilde{\mu}, \tilde{u})) \le \mathcal{V}_{min}((\mu, u))$. So it must be that $(\tilde{\mu}, \tilde{u}) \in \mathcal{L}_{min}$ and $(\tilde{\mu}, \tilde{u}) \in \mathcal{SS}$. □

Proof of Lemma 1. By statement of the Lemma, $(\mu, u) \in \mathcal{S}(W, F, v)$ is such that for all $j \neq \mu(j)$, $u_j + u_{\mu(j)} > 0$; $i \in \Delta_0, u_i > 0$. We show how the outcome of a 1-error by agent i can be reached via a 0-error.

If $u_{\mu(i)} = 0$, then let (μ^1, u^1) be obtained from (μ, u) by $\mu(i)$ becoming single. $u^1_i = u^1_{\mu(i)} = 0$. 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 obtained from (μ^1, u^1) by $(i, \mu(i))$ matching with payoffs $u^2_i = u_i - 1$ and $u^2_{\mu(i)} = u_{\mu(i)} + 1$. (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 2(a) we know that $\mu^*(\mu(i)) \neq \mu(i)$. By Lemma 2(c), $(\mu^*, u) \in \mathcal{S}(W, F, v)$ and $u_{\mu^*(\mu(i))} + u_{\mu(i)} = v(\mu^*(\mu(i)), \mu(i))$. Let (μ^1, u^1) be obtained from (μ, u) by $(\mu^*(\mu(i)), \mu(i))$ matching with payoffs $u^1_{\mu^*(\mu(i))} = u_{\mu^*(\mu(i))}$ and $u^1_{\mu(i)} = u_{\mu(i)}$. Now, $\mu^1(i) = i, u^1_i = 0$, so $u_{\mu(i)} > 0$ implies that $(i, \mu(i))$ is a blocking pair for (μ^1, u^1). Let (μ^2, u^2) be obtained from (μ^1, u^1) by $(i, \mu(i))$ matching, $u^2_i = u_i - 1, u^2_{\mu(i)} = u_{\mu(i)} + 1$. 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 obtained from (μ^2, u^2) by $(\mu^*(\mu(i)), \mu(\mu^*(\mu(i))))$ matching with payoffs $u^3_{\mu^*(\mu(i))} = u_{\mu^*(\mu(i))}$ and $u^3_{\mu(\mu^*(\mu(i)))} = u_{\mu(\mu^*(\mu(i)))}$. (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 $\tilde{\mu}'$. By Lemma 2(b), matchings μ, μ' are optimal and so are $\tilde{\mu}, \tilde{\mu}'$.

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 by Lemma 2(c), $(\mu^2, u) \in \mathcal{S}$. Iterating, we eventually obtain $(\mu^{L_1}, u) = (\mu', u) \in \mathcal{S}$. 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 $(o^{L_1}, \ldots, o^{L_2})$, $o^{L_1} = (\mu', u), o^{L_2} = (\tilde{\mu}', u) \in \mathcal{S}$, 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 1 can be replicated by a 0-error.

From $o^{L_2} = (\tilde{\mu}', u) \in \mathcal{S}$, 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 by Lemma 2(c), $(\mu^{L_2+1}, u) \in \mathcal{S}$. Iterating, we eventually obtain $o^{L_3} = (\tilde{\mu}, u) \in \mathcal{S}$.

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 payoff closer to $(\tilde{\mu}, \tilde{u})$ than $(\tilde{\mu}, u)$ is. Lemma 1 shows that such a 1-error can be replicated by a 0-error. Therefore, there exists a sequence of stable outcomes $(o^{L_3}, \ldots, o^{L_4})$, $o^{L_3} = (\tilde{\mu}, u), o^{L_4} = (\tilde{\mu}, u^{L_4}) \in \mathcal{S}, u^{L_4}_i = \tilde{u}_i$ 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. Rev. Econom. Stud. 64 (3), 411–426.

Agastya, M., 1999. Perturbed adaptive dynamics in coalition form games. J. Econom. Theory 89 (2), 207–233.

Biró, P., Bomhoff, M., Golovach, P.A., Kern, W., Paulusma, D., 2013. Solutions for the stable roommates problem with payments. In: Graph-Theoretic Concepts in Computer Science. In: Lecture Notes in Computer Science, Springer Verlag, pp. 69–80.

Blume, L.E., 1993. The statistical mechanics of strategic interaction. Games Econom. Behav. 5 (3), 387–424.

Chen, B., Fujishige, S., Yang, Z., 2012. Decentralized market processes to stable job matchings with competitive salaries. Working paper, Department of Economics, University of York.

Crawford, V.P., Knoer, E.M., 1981. Job matching with heterogeneous firms and workers. Econometrica 49 (2), 437–450.

Demange, G., Gale, D., 1985. The strategy structure of two-sided matching markets. Econometrica 53 (4), 873–888.

Diamantoudi, E., Xue, L., Miyagawa, E., 2004. Random paths to stability in the roommate problem. Games Econom. Behav. 48 (1), 18–28.

Dokumaci, E., Sandholm, W.H., 2011. Large deviations and multinomial probit choice. J. Econom. Theory 146 (5), 2151–2158.

Feldman, A.M., 1974. Recontracting stability. Econometrica 42 (1), 35–44.

Freidlin, M.I., Wentzell, A.D., 1984. Random Perturbations of Dynamical Systems, first ed.. Springer Verlag, Second edition 1998.

Green, J.R., 1974. The stability of Edgeworth's recontracting process. Econometrica 42 (1), 21–34.

Jackson, M.O., Watts, A., 2002. The evolution of social and economic networks. J. Econom. Theory 106 (2), 265–295.

Klaus, B., Klijn, F., 2007. Paths to stability for matching markets with couples. Games Econom. Behav. 36 (1), 154–171.

Klaus, B., Klijn, F., Walzl, M., 2010. Stochastic stability for roommate markets. J. Econom. Theory 145 (6), 2218–2240.

Klaus, B., Payot, F., 2015. Paths to stability in the assignment problem. J. Dyn. Games (forthcoming).

Kojima, F., Ünver, M.U., 2008. Random paths to pairwise stability in many-to-many matching problems: A study on market equilibration. Internat. J. Game Theory 36 (3–4), 473–488.

Lim, W., Neary, P.R., 2013. An Experimental Investigation of Stochastic Stability. Technical report. Hong Kong University of Science and Technology.

Mäs, M., Nax, H.H., 2014. A behavioral study of noise in coordination games. mimeo.

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

Myatt, D.P., Wallace, C., 2003. A multinomial probit model of stochastic evolution. J. Econom. Theory 113 (2), 286–301.

Myerson, R., 1978. Refinements of the Nash equilibrium concept. Internat. J. Game Theory 7, 73–80.

Nax, H.H., Pradelski, B.S.R., 2015. Evolutionary dynamics and equitable core selection in assignment games. Internat. J. Game Theory 44 (4), 903–932.

Nax, H.H., Pradelski, B.S.R., 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, pp. 2391–2397.

Newton, J., 2012a. Coalitional stochastic stability. Games Econom. Behav. 75 (2), 842–854.

Newton, J., 2012b. Recontracting and stochastic stability in cooperative games. J. Econom. Theory 147 (1), 364–381.

Newton, J., 2015. Stochastic stability on general state spaces. J. Math. Econom. 58, 46–60.

Newton, J., Sawa, R., 2015. A one-shot deviation principle for stability in matching problems. J. Econom. Theory 157, 1–27.

Nöldeke, G., Samuelson, L., 1993. An evolutionary analysis of backward and forward induction. Games Econom. Behav. 5 (3), 425–454.

Núñez, M., Rafels, C., 2008. On the dimension of the core of the assignment game. Games Econom. Behav. 64 (1), 290–302.

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

Roth, A.E., Vande Vate, J.H., 1990. Random paths to stability in two-sided matching. Econometrica 58 (6), 1475–1480.

Sawa, R., 2014. Coalitional stochastic stability in games, networks and markets. Games Econom. Behav. 88, 90–111.

Selten, R., 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. Internat. J. Game Theory 4 (1), 25–55.

Sengupta, A., Sengupta, K., 1996. A property of the core. Games Econom. Behav. 12 (2), 266–273.

Serrano, R., Volij, O., 2008. Mistakes in cooperation: The stochastic stability of edgeworth's recontracting. Econ. J. 118 (532), 1719–1741.

Shapley, L.S., Shubik, M., 1971. The assignment game I: The core. Internat. J. Game Theory 1 (1), 111–130.

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.