Paths to Stability in the Assignment Problem*

Bettina Klaus[†] Frédéric Payot[‡]

September 24, 2013

Abstract

We study a labor market with finitely many heterogeneous workers and firms to illustrate the decentralized (myopic) blocking dynamics in two-sided one-to-one matching markets with continuous side payments (assignment problems, Shapley and Shubik, 1971).

A labor market is unstable if there is at least one blocking pair, that is, a worker and a firm who would prefer to be matched to each other in order to obtain higher payoffs than the payoffs they obtain by being matched to their current partners. A blocking path is a sequence of outcomes (specifying matchings and payoffs) such that each outcome is obtained from the previous one by satisfying a blocking pair (i.e., by matching the two blocking agents and assigning new payoffs to them that are higher than the ones they received before).

We are interested in the question if starting from any (unstable) outcome, there always exists a blocking path that will lead to a stable outcome. In contrast to discrete versions of the model (i.e., for marriage markets, one-to-one matching, or discretized assignment problems), the existence of blocking paths to stability cannot always be guaranteed. We identify a necessary and sufficient condition for an assignment problem (the existence of a stable outcome such that all matched agents receive positive payoffs) to guarantee the existence of paths to stability and show how to construct such a path whenever this is possible.

JEL classification: C71, C78, D63.

Keywords: Assignment problem, competitive equilibria, core, decentralized market, random path, stability.

1 Introduction

Many markets involve bilateral relationships where each agent of one side of the market can be matched to any agent of the other side of the market but cannot be matched to any agent from the same side. Examples for such two-sided matching markets include marriage markets (women and men), college admissions markets (colleges and students), auction markets (buyers and sellers), and labor markets (workers and firms). Two-sided matching markets can be partitioned in two main categories: markets without side payments (e.g., marriage and college admissions markets) and markets with side payments (e.g., auction and labor markets). Side payments in the form

*We thank Mehmet Karakaya, Alexandru Nichifor, Jan Christoph Schlegel, and Markus Walzl for their valuable comments and interesting discussions.

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

‡The results of this article have been published as part of the PhD thesis of Frédéric Payot in 2011 (Payot, 2011, Chapter 3).

of prices, fees, or salaries are a natural feature of many economic situations. Here, we study a simple two-sided one-to-one matching market with side payments: a labor market with finitely many heterogeneous workers and firms. To keep the model simple we impose a unit-demand condition such that each worker accepts at most one job and each firm hires at most one worker.

Two-sided matching markets with side payments –*assignment problems*– have first been analyzed by Shapley and Shubik (1971). In an assignment problem, indivisible objects (e.g., auctioned items or jobs) are exchanged with monetary transfers (e.g., prices or salaries) between two finite sets of agents (e.g., buyers/sellers or workers/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 can form pairs to exchange the corresponding objects and at the same time make monetary transfers of the value they create (alternatively, singletons can execute an outside option).

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 stable if it is individually rational and there is no blocking pair, that is, there are no two agents that are not matched with each other, but in fact would prefer to be. For instance, in a labor market, a worker and a firm form a blocking pair if both could get higher payoffs than the payoffs they obtain by being matched to their current partners (if we matched them with higher payoffs, we would be satisfying a blocking par). An outcome is in the *core* if no coalition of agents can improve their payoffs by rematching among themselves. Shapley and Shubik (1971) showed that (a) the core of an assignment problem and the set of stable outcomes coincide, (b) for any assignment problem, there always exists a stable outcome, (c) the set of stable outcomes is a complete lattice with two extreme points, each of them corresponding to an outcome where all the agents of the same side of the market (e.g., the workers) receive their maximal stable payoffs while the agents of the other side (e.g., the firms) receive their minimal stable payoffs, and (d) at any stable outcome, the matching between the workers and the firms is optimal (i.e., the value created by the pairs in the corresponding matching is maximal). Sotomayor (2003) and Wako (2006) proved that if there is only one optimal matching, then the core contains infinitely many stable outcomes. Conversely, the core is a singleton only when multiple optimal matchings exist.

The literature on stability in two-sided matching markets was initiated by Gale and Shapley (1962) who proposed a centralized procedure, the famous deferred acceptance algorithm, to find a stable outcome for any marriage or college admissions problem (with responsive preferences). The deferred acceptance algorithm turned out to be the key element for many centralized market clearing houses, e.g., for the National Resident Matching Program (Roth, 1984), for school choice programs (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2005), and for auctions and trading networks (Demange et al., 1986; Gul and Stacchetti, 2000; Milgrom, 2000; Ausubel, 2006; Sun and Yang, 2009; Hatfield et al., 2013).[1]

Dynamic changes in real world (two-sided) matching markets are frequently observed. This indicates that outcomes often are not stable. For instance, in a labor market, a worker might

[1]See also Crawford and Knoer (1981), Kelso and Crawford (1982), and Crawford (2008) for centralized processes in labor markets and Roth and Sotomayor (1990) for an excellent survey on two-sided matching theory until 1990.

switch to a new job if that increases his salary while the firm who hires him finds his qualification/productivity higher than that of his predecessor. A blocking path for a (two-sided) matching markets is a finite sequence of outcomes where each outcome is obtained from the previous one by satisfying a blocking pair taking into account that agents behave myopically, i.e., agents do not forecast how their decision to block an outcome might influence the future evolution of the market.

Knuth (1976) showed that for marriage markets a process of myopic blocking may cycle, i.e., a decentralized process may not converge to a stable outcome. Roth and Vande Vate (1990) show that for marriage markets there always exists a blocking path starting from any unstable outcome that leads to a stable outcome in finitely many steps. Assuming that each blocking pair is selected with strict positive probability, this result implies that a decentralized blocking process converges to stability with probability one.[2] Chen et al. (2012) and Nax et al. (2013) both analyze a similar decentralized blocking process for labor markets with discrete side payments. As in Roth and Vande Vate (1990), they construct a blocking path to stability and show that a decentralized blocking process converges to stability with probability one. Biró et al. (2013) consider a more general one-sided version of the assignment problem and obtain results that imply those of Chen et al. (2012) and Nax et al. (2013) with a different proof technique. Apart from looking at a continuous model instead of a discretized one (as Chen et al. (2012) and Nax et al. (2013) do), a difference between the work of Biró et al. (2013), Chen et al. (2012), Nax et al. (2013) compared to ours is that we consider strict blocking while all these other articles consider weak blocking. This difference induces differentiated results and different proof techniques and we will discuss the exact relation of these articles with ours in Section 4.3.

The paper is organized as follows:

In Section 2, we introduce the classical assignment model with continuous side payments (Shapley and Shubik, 1971).

In Section 3, we define a generic blocking path and we show with a few examples that a blocking path to stability might not exist for all assignment problems. We then state and prove our main result that, for all assignment problems that satisfy our necessary and sufficient condition (the existence of a stable outcome such that all matched agents receive positive payoffs), a stable outcome can always be obtained through a finite sequence of outcomes, each outcome being obtained from the previous one by satisfying a blocking pair under the strict blocking norm.

Finally, in Section 4, we discuss some relevant points. First, we consider a specific blocking path where each time a blocking pair is satisfied the blocking agents equally split the surplus they create. We ask whether such a fair blocking path can be always used to construct a path to stability (the answer is no). Second, we discuss the probabilistic interpretation of the blocking path result obtained in Section 3. Third, we discuss in more details the articles by Biró et al. (2013), Chen et al. (2012), and Nax et al. (2013), and show how their results and our results are related. Fourth, we propose a centralized stabilization procedure that uses a so-called median stable outcome as compromise target outcome. Then, we briefly conclude.

[2]Diamantoudi et al. (2004) establish a similar result for (one-to-one matching) roommate problems and Klaus and Klijn (2007) for matching markets with couples.

2 The Assignment Problem

We consider a simple labor market model that matches workers and firms. 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.[3] We denote the set of pairs that agents in $W \times F$ can form (including "degenerate" pairs where agents $i \in W \cup F$ form a "pair" (i,i) with themselves) by $P(W,F) = \{(w,f) \in W \times F\} \cup \{(i,i) \mid i \in W \cup F\}$.

A *characteristic function* for $W \cup F$ is a function $\pi : P(W,F) \to \mathbb{R}_+$ such that for each $i \in W \cup F$, $\pi(i,i) = 0$. The characteristic function π describes the *value* that agents create when forming pairs. In particular, $\pi(i,i) = 0$ represents the *reservation value* of an agent $i \in W \cup F$.[4] A *(two-sided one-to-one) assignment problem* is a triple (W, F, π).

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

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

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

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

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

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

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 an optimal matching μ such that $(w,f) \in \mu$.

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

---

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

[4]Our assumptions on the characteristic function π are without loss of generality. 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 $\pi(w,f) = \max\{(s_{\max}(w,f) - s_{\min}(w,f)), 0\} \geq 0$.

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

(ii) if $(i, i) \in \mu$, then $u_i = \pi(i, i) = 0$.

The following property is a voluntary participation condition based on the idea that an agent can always enforce his reservation value by staying single. An outcome (μ, u) [a payoff vector u] is *individually rational* if for each $i \in W \cup F$, $u_i \geq 0$.

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

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

An outcome (μ, u) [a payoff vector u] is *stable* if

(a) it is individually rational, i.e., for all $i \in W \cup F$, $u_i \geq 0$ and

(b) no blocking pairs exist, i.e., for all $(w, f) \in W \times F$, $u_w + u_f \geq \pi(w, f)$.[5]

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

Remark 1 (Stable Outcomes)**.** The set of stable outcomes coincides with the core (Shapley and Shubik, 1971): we could model an assignment problem (W, F, π) as a cooperative game with transferable utility (TU) whose characteristic function v assigns to each coalition $S \subseteq W \cup F$, the number $v(S) \equiv \max_{\mu \in \mathcal{M}(W \cap S, F \cap S)} \{\sum_{(i,j) \in \mu} \pi(w, f)\}$ with $v(\emptyset) = 0$. The core of assignment problem (W, F, π) is the set $\mathcal{C}(W, F, \pi) = \{(\mu, u) \mid \mu$ is optimal and for all $S \subseteq W \cup F, \sum_{i \in S} u_i \geq v(S)\}$. Thus, for any assignment problem (W, F, π) an outcome (μ, u) is in the core if matching μ is optimal and no coalition of agents $S \subseteq W \cup F$ can improve their payoffs at u by rematching among themselves. Furthermore, if an agent is single at a stable outcome, then at each stable outcome, he receives his reservation value (Demange and Gale, 1985). $\triangle$

The following lemma explains how optimal matchings and stable payoffs are related.

Lemma 1 (**Optimal matchings and stable outcomes**)**.**
(a) *If* (μ, u) *is a stable outcome for assignment problem* (W, F, π)*, then* μ *is an optimal matching for assignment problem* (W, F, π) (Roth and Sotomayor, 1990, Corollary 8.8)*.*
(b) *Let* (μ, u) *be a stable outcome and* μ' *be an optimal matching for assignment problem* (W, F, π)*. Then,* (μ', u) *is a stable outcome for assignment problem* (W, F, π) (Roth and Sotomayor, 1990, Corollary 8.7)*.*

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

[5] Note that in the definition of an outcome we could replace conditions (i) and (ii) by $[\sum_{i \in W \cup F} u_i = \sum_{(i,j) \in \mu} \pi(i, j)]$. Then, a stable outcome (μ, u) would automatically satisfy (i) and (ii) in the definition of an outcome (Roth and Sotomayor, 1990, Lemma 8.5). Defining an outcome via conditions (i) and (ii) simplifies our exposition, but it is not essential for our results.

Let (W, F, π) be an assignment problem. Then, for each agent $i \in W \cup F$ the *set of stable payoffs* equals a closed interval

$$[\underline{u}_i, \overline{u}_i] = \{u'_i \in \mathbb{R}_+ \mid \text{there exists a stable outcome } (\mu, u) \text{ such that } u_i = u'_i\}.$$

Thus, $\underline{u}_i$ is the *minimal stable payoff of agent* i and $\overline{u}_i$ is the *maximal stable payoff of agent* i. Let $\underline{u}_W := (\underline{u}_w)_{w\in W}$, $\underline{u}_F := (\underline{u}_f)_{f\in F}$, $\overline{u}_W := (\overline{u}_w)_{w\in W}$, and $\overline{u}_F := (\overline{u}_f)_{f\in F}$. If, at some arbitrary outcome (μ, u), agent i receives a payoff $u_i \notin [\underline{u}_i, \overline{u}_i]$, then we say that agent i receives an *unstable payoff*.

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

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

We provide two examples to illustrate the set of stable outcomes and their properties.

Example 1 (An infinite set of stable outcomes). Let (W, F, π) be an assignment problem given by $W = \{w\}$, $F = \{f\}$, and $\pi(w, f) = 1$; that is, worker w and firm f generate value 1 by forming a pair. In any stable outcome (μ, u), w and f are matched and the set of stable outcomes equals $\mathcal{S}(W, F, \pi) := \{(\mu, u) \mid (w, f) \in \mu \text{ and } [u_w \geq 0, u_f \geq 0, \text{ and } u_w + u_f = 1]\}$. Note that there exists a unique optimal matching μ and that $[\underline{u}_w, \overline{u}_w] = [\underline{u}_f, \overline{u}_f] = [0, 1]$, $(\overline{u}_W, \underline{u}_F) = (1, 0)$, and $(\underline{u}_W, \overline{u}_F) = (0, 1)$. △

Example 2 (A finite set of stable outcomes). Let (W, F, π) be an assignment problem given by $W = \{w_1, w_2\}$, $F = \{f\}$, and $\pi(w_1, f) = \pi(w_2, f) = 1$. A stable outcome (μ, u) for (W, F, π) matches one of the workers $w \in W$ with firm f while the other worker is single and the firm obtains the total value. Formally, the set of stable outcomes is $\mathcal{S}(W, F, \pi) := \{(\mu, u) \mid [(w_1, f) \in \mu \text{ or } (w_2, f) \in \mu] \text{ and } [u_{w_1} = u_{w_2} = 0 \text{ and } u_f = 1]\}$. Any outcome (μ, u) at which firm f earns $u_f < 1$ is not stable because then there is always a single worker w with $u_w = 0$ such that $u_w + u_f < 1 = \pi(w, f)$. Note that there exist two optimal matchings and that $[\underline{u}_{w_1}, \overline{u}_{w_1}] = [\underline{u}_{w_2}, \overline{u}_{w_2}] = \{0\}$, $[\underline{u}_f, \overline{u}_f] = \{1\}$, and $(\overline{u}_W, \underline{u}_F) = (\underline{u}_W, \overline{u}_F) = (0, 0, 1)$. △

In Example 1 the set of stable outcomes is infinite: a unique optimal matching supports an infinite number of stable payoffs. In contrast, the set of stable outcomes in Example 2 is finite: two optimal matchings support the unique stable payoff.

3 Blocking Paths to Stability

For the following definitions of paths and blocking paths and in the formulation of our results we only focus on non-degenerate blocking pairs (of one worker with one firm). Our focus on non-degenerate blocking pairs eases notation without loss of generality since we can always obtain an individually rational outcome whenever needed by "singleton blocking" by those agents who obtain negative payoffs.

A *path* for assignment problem (W,F,π) is a (finite!) sequence of outcomes $(\mu^1,u^1),...,(\mu^k,u^k)$ such that for each $l \in \{1,...,k-1\}$, the outcome (μ^{l+1},u^{l+1}) is obtained from (μ^l,u^l) by matching a pair (w_l,f_l). This induces the matching μ^{l+1}

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

and the payoff vector u^{l+1}

$$u^{l+1}_x=\begin{cases} u^{l+1}_{w_l} & \text{if } x=w_l\\ u^{l+1}_{f_l} & \text{if } x=f_l\\ 0 & \text{if } x\neq w_l,f_l \text{ and } x\in\{\mu^l(w_l),\mu^l(f_l)\}\\ u^l_x & \text{otherwise}\end{cases}$$

such that $u^{l+1}_{w_l}+u^{l+1}_{f_l}=\pi(w_l,f_l)$. Thus, at outcome (μ^{l+1},u^{l+1}), agents w_l and f_l are matched and generate value $\pi(w_l,f_l)$, their former partners 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,π) is a path of individually rational outcomes $(\mu^1,u^1),...,(\mu^k,u^k)$ such that for each $l\in\{1,...,k-1\}$, the outcome (μ^{l+1},u^{l+1}) is obtained from (μ^l,u^l) by matching a blocking pair (w_l,f_l) for (μ^l,u^l) such that the corresponding payoffs are $u^{l+1}_{w_l}>u^l_{w_l}$ and $u^{l+1}_{f_l}>u^l_{f_l}$, i.e., the blocking agents w_l and f_l split their blocking surplus such that each of them is strictly better off at outcome (μ^{l+1},u^{l+1}). Hence, while Chen et al. (2012), Nax et al. (2013), and Biró et al. (2013) require weak blocking,[6] we require the more demanding *strict blocking* norm for blocking pairs. We say that a blocking path *leads to stability* if the last outcome (μ^k,u^k) is stable. We give a simple illustration using the assignment problem introduced in Example 1.

Example 3 (**A blocking path to stability**)**.** Consider the assignment problem (W,F,π) in Example 1: $W=\{w\}$, $F=\{f\}$, and $\pi(w,f)=1$. Start the sequence with the empty matching (μ^1,u^1) as the initial (unstable) outcome, i.e., $\mu^1(w)=w,\mu^1(f)=f$, and $u^1_w=u^1_f=0$. Note that if w and f form a pair, then their blocking surplus equals $bs(u^1;(w,f))=1$. Let (μ^2,u^2) be obtained from (μ^1,u^1) by satisfying this blocking pair using an equal split of the blocking surplus, i.e., $u^2_w=u^2_f=\frac{1}{2}$. Then, (μ^2,u^2) is stable and the blocking path $(\mu^1,u^1),(\mu^2,u^2)$ leads to stability in one step. △

Example 3 only shows that such a blocking path to stability might exist. In the next example we construct an infinite sequence of outcomes by satisfying blocking pairs. Recall that a blocking path to stability is a finite sequence of outcomes ending in a stable outcome; hence, the following example does not construct a path to stability. In fact, we prove in Theorem 2 that no path to stability exists for the following two examples (Examples 4 and 5).

[6] Under the *weak blocking* norm, agents payoffs only need to satisfy $u^{l+1}_{w_l}\geq u^l_{w_l}$ and $u^{l+1}_{f_l}\geq u^l_{f_l}$ with at least one strict inequality.

Example 4 (**An infinite sequence converging to stable payoffs**). Consider the assignment problem (W, F, π) in Example 2: $W = \{w_1, w_2\}$, $F = \{f\}$, and $\pi(w_1, f) = \pi(w_2, f) = 1$. We construct a sequence $(\mu^1, u^1), (\mu^2, u^2), (\mu^3, u^3), \ldots$, such that each outcome is obtained by matching a blocking pair with the additional property that the blocking pair equally splits the blocking surplus. Consider outcome (μ^l, u^l) where firm f has a payoff $u^l_f < 1$. The blocking surplus of f with the single worker w_s is $bs(u^l; (w_s, f)) = 1 - u^l_f - 0$. Hence, when equally splitting the blocking surplus, we obtain $u^{l+1}_f = u^l_f + \frac{1}{2}(1 - u^l_f)$ and $u^{l+1}_{w_s} = \frac{1}{2}(1 - u^l_f)$.

Start the sequence with the empty matching (μ^1, u^1) as the initial (unstable) outcome, i.e., all agents are single and receive zero payoffs. Select worker w_1 and let (μ^2, u^2) be the outcome obtained from (μ^1, u^1) by satisfying the blocking pair (w_1, f), such that $\mu^2(f) = w_1, u^2_{w_1} = \frac{1}{2}, u^2_{w_2} = 0$, and $u^2_f = \frac{1}{2}$. Now, (w_2, f) is a blocking pair for (μ^2, u^2). Satisfy (w_2, f) to obtain the next outcome (μ^3, u^3), and so on. The following table summarizes the payoffs along the sequence.

l	1	2	3	4	...	k
$u^l_{w_1}$	0	$\frac{1}{2}$	0	$\frac{1}{8}$	...	$\begin{cases} 1 - u^k_f & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases}$
$u^l_{w_2}$	0	0	$\frac{1}{4}$	0	...	$\begin{cases} 0 & \text{if } k \text{ is even} \\ 1 - u^k_f & \text{if } k \text{ is odd} \end{cases}$
u^l_f	0	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{7}{8}$	...	$u^k_f = \sum_{i=1}^{k-1} (\frac{1}{2})^i = 1 - (\frac{1}{2})^{k-1}$

At each outcome (μ^l, u^l), if l is even, then firm f is matched to worker w_1 and if l is odd (except for $l = 1$), then firm f is matched to worker w_2 . Since for all $l \geq 1, u^l_f < 1$, the outcome (μ^l, u^l) is never stable. Therefore, the sequence $(\mu^1, u^1), (\mu^2, u^2), (\mu^3, u^3), \ldots$ is infinite. Furthermore, $\lim_{k\to\infty} u^k_f = 1$ and, for all $w \in W$, $\lim_{k\to\infty} u^k_w = 0$, i.e., payoffs converge to the unique stable payoffs. △

Example 4 shows that a blocking path to stability might not always exist (since the finiteness of such a path for this specific assignment problem is impossible - a statement we will proof formally when proving Theorem 2). In this example, that is the case because it is always possible to make the firm and the respective single worker better off by letting them block the previous outcome. The only way for a path in this example to end in a stable outcome would require the firm to obtain the total value generated by the blocking pair. But then, the worker that blocks with the firm would be indifferent between working for the firm and staying single (he receives zero payoff in both cases) – it is part of the definition of a blocking path that both agents in a blocking pair are better off.

Example 4 also illustrates that an infinite blocking sequence can converge to a stable payoff: the firm's payoff monotonically increases and converges towards 1 (its stable payoff) and the workers' payoffs converge to 0 (their stable payoffs) along the sequence. However, next we show that we cannot always guarantee convergence to stable payoffs: we vary the previous example by constructing an infinite sequence of outcomes that converge to unstable payoffs (in fact, workers' payoffs do not converge at all).

Example 5 (An infinite sequence not converging to stable payoffs). Let $a \in (0,1)$ and consider the assignment problem (W, F, π) in Examples 2 and 4: $W = \{w_1, w_2\}$, $F = \{f\}$, and $\pi(w_1, f) = \pi(w_2, f) = 1$. Unlike in Example 4, we construct a sequence that is not based on equally splitting the blocking surplus. Instead, if at outcome (μ^l, u^l) firm f has a payoff $u^l_f < a$, then its payoff at the next outcome by blocking with the single worker w_s is given by $u^{l+1}_f = u^l_f + \frac{1}{2}(a - u^l_f)$ and w_s's payoff is $u^{l+1}_{w_s} = (1-a) + \frac{1}{2}(a - u^l_f)$, i.e., we guarantee the single worker a minimal payoff $(1-a)$ whenever he blocks with firm f.

Similarly to Example 4, start the sequence with the empty matching (μ^1, u^1) as the initial (unstable) outcome. Select worker w_1 and let (μ^2, u^2) be the outcome obtained from (μ^1, u^1) by satisfying the blocking pair (w_1, f), such that $\mu^2(f) = w_1, u^2_{w_1} = 1 - \frac{1}{2}a, u^2_{w_2} = 0$, and $u^2_f = \frac{1}{2}a$. Continue the construction of the sequence similarly as in the previous example. The following table summarizes the payoffs along the sequence.

l	1	2	3	4	...	k
$u^l_{w_1}$	0	$1 - \frac{1}{2}a$	0	$1 - \frac{7}{8}a$	...	$\begin{cases} 1 - u^k_f & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{cases}$
$u^l_{w_2}$	0	0	$1 - \frac{3}{4}a$	0	...	$\begin{cases} 0 & \text{if } k \text{ is even} \\ 1 - u^k_f & \text{if } k \text{ is odd} \end{cases}$
u^l_f	0	$\frac{1}{2}a$	$\frac{3}{4}a$	$\frac{7}{8}a$	...	$u^k_f = \sum_{i=1}^{l-1} (\frac{1}{2})^i a = \left(1 - (\frac{1}{2})^{l-1}\right) a$

As in Example 4, f alternates between w_1 and w_2 as its partner. Since for all $l \geq 1$, $u^l_f < a < 1$, the outcome (μ^l, u^l) is never stable. Therefore, the sequence $(\mu^1, u^1), (\mu^2, u^2), (\mu^3, u^3), ...$ is infinite. Furthermore, and in contrast to Example 4, $\lim_{k \to \infty} u^k_f = a < 1$ and for workers w payoffs alternate between 0 and values $> (1-a)$, i.e., the firm's payoffs converge to an unstable payoff and workers payoffs do not converge. △

Note that in Examples 4 and 5, although the payoffs (partially) converge, the matchings do not converge. Along the blocking path firm f alternates between w_1 and w_2 as its partner. Because the path is not finite (and payoffs only converge in the limit, if at all) this oscillation never stops.

Example 2 is an assignment problem with a finite set of stable outcomes because the set of stable payoffs is a singleton. Based on this assignment problem, we have constructed in Examples 4 and 5 infinite sequences of outcomes that either payoff-converge to stability or that diverge in payoffs. In contrast, the set of stable outcomes in Example 1 is infinite because the set of stable payoffs is infinite. Based on the assignment problem described in Example 1 we have shown in Example 3 that a blocking path to stability might exist.

A crucial difference between the assignments problems depicted in Example 1 and Example 2 relates to the characteristics of the stable payoffs. In Example 1, there is an infinite number of *interior* stable outcomes, i.e., stable outcomes where all the agents obtain strictly positive payoffs, and two extreme outcomes where one of the two matched agents obtains a zero payoff. In Example 2, all the workers obtain zero payoffs at a stable outcome irrespective of which one of the two workers is matched with the firm. Since we require any two blocking agents to be strictly better off when satisfying a blocking pair, stability in Example 4 (based on Example 2) will never be reached because it is impossible to satisfy a blocking pair formed by a single worker and the firm with stable payoffs that make both of them strictly better off.[7] We show that for an assignment problem the existence of a stable outcome that is "away from zero" (zero being the normalized reservation value) for matched agents is a necessary and sufficient condition to guarantee the existence of a blocking path to stability. We formalize this condition as follows.

Assumption 1. *Assignment problem* (W, F, π) *satisfies Assumption 1 if there exists a stable outcome* (μ^*, u^*) *such that for each agent* $i \in W \cup F$ *who is not single, i.e.,* $\mu^*(i) \neq i$*, we have* $u_i^* > 0$.

Next, we show that Assumption 1 is sufficient for the existence of blocking paths to stability.

Theorem 1 (Paths to Stability)**.** *Let* (W, F, π) *be an assignment problem satisfying Assumption 1 and* (μ, u) *an arbitrary outcome for* (W, F, π)*. Then, there exists a blocking path* $(\mu^1, u^1), ..., (\mu^k, u^k)$ *such that* $(\mu, u) = (\mu^1, u^1)$ *and* (μ^k, u^k) *is stable.*

The proof of the theorem proceeds in three steps. In Step 1 we first unmatch as many couples as possible via blocking, i.e., we maximize the number of single agents. In some cases, using Assumption 1, we might already be able to rematch single agents via blocking to obtain a stable outcome. If this is not immediately possible, in Step 2 we then apply a stabilization process that deals with non optimal matchings, unstable payoffs, and single agents who cannot immediately be matched via blocking because they would have to receive a stable zero payoff to reach stability directly, which is not possible according to our strict blocking norm (and hence, some extra steps are needed to move the process towards a positive stable payoff for such agents). We prove that the stabilization process ends in finitely many steps by induction. In the final Step 3, using Assumption 1, we complete the construction of the blocking path by matching remaining single agents with optimal partners. The result is a stable outcome. Throughout the proof, a stable outcome (μ^*, u^*) as specified in Assumption 1 serves as a target and outcomes along the blocking path are getting closer to the target outcome along the way. We prove Theorem 1 in Appendix A and provide an example that illustrates the paths to stability construction in Appendix B.

Remark 2 (Induced and Related Results)**.**
Chen et al. (2012, Theorem 1) and Nax et al. (2013, Theorem 1) both analyze a similar decentralized (weak) blocking process for assignment problems with discrete side payments. They construct a blocking path to stability and show that a decentralized blocking process converges to stability with probability one.

[7] We shall discuss at the end of the paper the case of weak blocking, i.e., the relaxation of the strict blocking norm that only requires that both blocking agents are weakly better off and *at least* one of them strictly better off.

We show in Section 4.2 that in our continuous model, we cannot obtain a similar probabilistic interpretation of our paths to stability result. A probabilistic result however can easily be restored if we use the notion of ϵ-stability we discuss in Section 4.2.

In Section 4.3 we discuss how our result does imply the paths to stability results established by Chen et al. (2012) and Nax et al. (2013) but with a different proof technique that also has to address complications due to the use of our strict instead of the weak blocking norm. Our proof technique involves a stable target outcome instead of using the polarization that is inherent in the core for two-sided assignment problems.

This proof technique makes it possible to extend our result (Theorem 1) to one-sided assignment problems. Biró et al. (2013, Theorem 1) analyze decentralized (weak) blocking process for one-sided assignment problems and we discuss the relationship between their paths to stability result and ours in Section 4.3. Related to the target proof technique we and Biró et al. (2013) use, in Section 4.4 we discuss the new idea of using median stable target outcomes in a centralized stabilization process in order to implement a compromise in the final stable outcome. △

One distinguishing difference between our results (Theorems 1 and 2) and the results by Biró et al. (2013, Theorem 1), Chen et al. (2012, Theorem 1), and Nax et al. (2013, Theorem 1) is that with the strict blocking norm and continuity, paths to stability do not always exist (they only do exist if and only if Assumption 1 is satisfied).

Theorem 2 (No Path to Stability)**.** *Let (W, F, π) be an assignment problem violating Assumption 1. Then, there exists an outcome (μ, u) for (W, F, π) such that no blocking path leads to stability.*

We prove Theorem 2 in Appendix A.

Table 1: Survey of the existence of paths to stability results for different assignment model specifications. Note that a third dimension we could add is if results also apply to the corresponding one-sided model extensions (Biró et al. (2013) prove their results for one-sided assignment problems and we can extend our results to one-sided assignments problems as well).

assignment problems	weak blocking	strict blocking
discrete	Chen et al. (2012) & Nax et al. (2013)	by Theorem 1 (see Section 4.3)
continuous	Biró et al. (2013)	Theorems 1 & 2 (iff Assumption 1)

4 Discussion

4.1 Fair Blocking Paths

We have targeted throughout the proof of Theorem 1 an outcome (μ^*, u^*) "away from zero" for matched agents (according to Assumption 1). In some steps of our blocking path we had to align payoffs according to the stable payoff vector u^*; in other words, at times we have used very specific payoff splits for certain blocking pairs. It is a natural question to ask if our result could also be obtained via an equal division blocking dynamics (as used in Example 4).

We call a *fair blocking path* a sequence of outcomes, such that each outcome is obtained from the previous one by satisfying a blocking pair with the additional condition that the blocking agents equally split the blocking surplus. The following example shows that a fair blocking path might not lead to stability.

Example 6 (A fair blocking path with an infinite sequence of outcomes). Let (W, F, π) be an assignment problem given by $W = \{w_1, w_2\}$, $F = \{f_1, f_2\}$, and for all $(w, f) \in W \times F$, $\pi(w, f) = 1$. A stable outcome (μ, u) for (W, F, π) matches each worker with any of the two firms and both workers (respectively both firms) obtain the same payoffs, i.e., stable payoffs must be aligned. Formally, the set of stable outcomes is $\mathcal{S}(W, F, \pi) := \{(\mu, u) \mid [(w_1, f_1), (w_2, f_2) \in \mu \text{ or } (w_1, f_2), (w_2, f_1) \in \mu] \text{ and } [u_{w_1} = u_{w_2} \text{ and } u_{f_1} = u_{f_2} = 1 - u_{w_1}]\}$. Any outcome (μ, u) at which the two workers obtain different payoffs is not stable because then there is always a worker w_i, $i \in \{1, 2\}$, with $u_{w_i} < u_{w_j}$, $i \neq j$, such that $u_{\mu(w_j)} < u_{\mu(w_i)}$ implies $u_{w_i} + u_{\mu(w_j)} < 1 = u_{w_i} + u_{\mu(w_i)}$. Thus, the worker who gets the smallest payoff always forms a blocking pair with the firm that is matched with the other worker.

We show that no fair blocking sequence $(\mu^1, u^1), (\mu^2, u^2), (\mu^3, u^3), \ldots$ can converge to stability. Start the sequence with outcome (μ^1, u^1) where worker w_1 is matched with firm f_1 and has a payoff $u^1_{w_1} = a \notin \{0, \frac{1}{2}, 1\}$, and worker w_2 and firm f_2 are single and obtain zero payoffs. Graphically, (μ^1, u^1) can be represented as follows:

$$(\mu^1, u^1) \qquad \begin{array}{ccc} w_1 & \xrightarrow{\quad a \quad}\!\!\xleftarrow{\quad 1-a \quad} & f_1 \\ w_2. & & .f_2 \end{array}$$

There are three blocking pairs for (μ^1, u^1): (w_1, f_2), (w_2, f_1), and (w_2, f_2), such that (μ^2, u^2) is obtained from (μ^1, u^1) by satisfying a blocking pair $(w, f) \in \{(w_1, f_2), (w_2, f_1), (w_2, f_2)\}$. We show that irrespective of which blocking pair is satisfied, the next outcome (μ^2, u^2) has the same unstable structure as outcome (μ^1, u^1), etc. We consider two cases: either (w, f) involves a single agent and a matched agent, i.e., $(w, f) \in \{(w_1, f_2), (w_2, f_1)\}$, or (w, f) involves two single agents, i.e., $(w, f) = (w_2, f_2)$.

Case 1 ($(w, f) \in \{(w_1, f_2), (w_2, f_1)\}$). The blocking surplus that w and f create is

$$bs(u^1; (w, f)) = \begin{cases} 1 - a & \text{if } (w, f) = (w_1, f_2) \\ a & \text{if } (w, f) = (w_2, f_1) \end{cases}$$

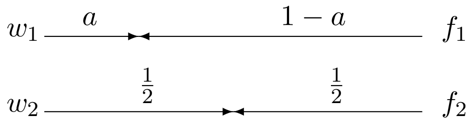
Since $a \notin \{0, 1\}$, the blocking surplus $bs(u^1; (w, f))$ is smaller than 1 irrespective of which blocking pair (w_1, f_2) or (w_2, f_1) is satisfied. Hence, at outcome (μ^2, u^2), w and f are matched, agent $i \in \{w, f\}$ who was single at μ^1 obtains a payoff $u^2_i = \frac{1}{2} bs(u^1; (w, f)) < \frac{1}{2}$, his partner $\mu^2(i)$ obtains a payoff $u^2_{\mu^2(i)} = 1 - u^2_i > \frac{1}{2}$, and their former partners $\mu^1(w)$ and $\mu^1(f)$ are single and obtain zero payoffs. Note that outcome (μ^2, u^2) has the same structure as outcome (μ^1, u^1), that is two agents w and f are matched and both of them obtain payoffs $u^2_w, u^2_f \notin \{0, \frac{1}{2}, 1\}$, and the two remaining agents are single and obtain zero payoffs.

Case 2 $((w, f) = (w_2, f_2))$**.** Since both w_2 and f_2 obtain zero payoffs at (μ^1, u^1) the blocking surplus they create is

$$bs(u^1; (w_2, f_2)) = 1$$

Satisfying this blocking pair leads to outcome (μ^2, u^2) where w_1 and f_1 are still matched with each other and obtain the same payoffs as before, and w_2 and f_2 are matched with each other and obtain payoffs $u^2_{w_2} = u^2_{f_2} = \frac{1}{2}$. Graphically, (μ^2, u^2) can be represented as follows:

(μ^2, u^2)



Since $a \notin \{0, \frac{1}{2}, 1\}$ by assumption, either w_1 or f_1 obtains a payoffs smaller than $\frac{1}{2}$ at (μ^2, u^2). Let $i \in \{w_1, f_1\}$ be the agent whose payoff at (μ^2, u^2) is smaller than $\frac{1}{2}$. Then there exists a matched agent $j \in \{w_2, f_2\}$, such that (i, j) is a blocking pair for (μ^2, u^2) with the blocking surplus

$$bs(u^2; (i, j)) < 1.$$

Note that (i, j) is the unique blocking pair for (μ^2, u^2). Satisfy this blocking pair with equal surplus splitting to obtain the next outcome (μ^3, u^3). At outcome (μ^3, u^3), j obtains a payoff $u^3_j = \frac{1}{2} + \frac{1}{2} bs(u^2; (i, j)) \notin \{0, \frac{1}{2}, 1\}$, which implies $u^3_i \notin \{0, \frac{1}{2}, 1\}$, and the two remaining agents are single and obtain zero payoffs. Hence, outcome (μ^3, u^3) has the same structure as outcome (μ^1, u^1).

Hence, in this example, any blocking path that starts from the unstable outcome (μ^1, u^1) at some point always reaches an unstable outcome that has the same structure as outcome (μ^1, u^1). Hence, the sequence $(\mu^1, u^1), (\mu^2, u^2), (\mu^3, u^3), \ldots$ is infinite and a fair blocking path in this example cannot lead to stability. $\triangle$

We have assumed in Example 6 that $u^1_{w_1} \notin \{0, \frac{1}{2}, 1\}$ and we have constructed an infinite sequence of outcomes. Of course, a fair blocking path might lead to stability in finitely many steps in some instances. Consider the following sequence: let (μ^1, u^1) be an outcome such that worker w_1 is matched with firm f_1 and both of them have payoffs $u^1_{w_1} = u^1_{f_1} = \frac{1}{2}$, and worker w_2 and firm f_2 are single and obtain zero payoffs. Suppose the next outcome (μ^2, u^2) is obtained from (μ^1, u^1) by satisfying the blocking pair (w_2, f_2) with equal surplus splitting. At outcome (μ^2, u^2), worker w_1 is matched with firm f_1, worker w_2 is matched with firm f_2 and all the agents obtain stable payoffs $u^2_i = \frac{1}{2}$, $i \in W \cup F$. Hence, the fair blocking path $(\mu^1, u^1), (\mu^2, u^2)$ leads to stability in one step.

4.2 Probabilistic Interpretation and ϵ-Stability

A central question addressed by Roth and Vande Vate (1990), Diamantoudi et al. (2004), Chen et al. (2012, Theorem 1), and Nax et al. (2013, Theorem 1) is whether a decentralized process where each blocking pair (and each possible blocking surplus split in the latter models) is randomly selected with a strictly positive probability converges to a stable outcome. All four papers answer this question in the affirmative. In each of these papers the authors construct a blocking path that leads to stability in finitely many steps. Since each blocking pair is selected with strictly positive probability in a decentralized process, the blocking path they construct converges to stability with probability one. However, the fact that each blocking pair is selected with positive probability relies precisely on the underlying assumptions of the models. In the marriage problem of Roth and Vande Vate (1990) and the roommate problem of Diamantoudi et al. (2004) agents have ordinal preferences over the (finite) set of agents with whom they can form a blocking pair. In the assignment problem of Chen et al. (2012) and Nax et al. (2013) side payments are discrete, such that the number of possible divisions of a blocking surplus is finite. Those assumptions imply that for each outcome there is always a finite number of blocking pairs (including no blocking pairs if the outcome is stable).

In our assignment problem with continuous side payments two blocking agents can split the blocking surplus in infinitely many ways. We replace the assumption of the above discrete models that any blocking pair and surplus split is chosen with a positive probability with the assumption that blocking pairs and surplus splits are based on a probability distribution with full support over all blocking pairs and surplus splits. Now, the probabilistic interpretation based on the existence of a path to stability turns out to be problematic in our model if in some step of our blocking path constructed stable payoffs have to be aligned (see our next example for such a situation). More precisely, if for some blocking pair there is a unique division of a blocking surplus that leads to stability, then, given the continuity of payoffs, the point probability that such a blocking pair is selected is zero. Hence, in our model, we cannot deduct a probabilistic convergence to stability result from the existence of a blocking path to stability. The following example illustrates the situation.

Example 7 (**Probabilistic Interpretation**)**.** Consider the assignment problem (W, F, π) in Example 6: $W = \{w_1, w_2\}$, $F = \{f_1, f_2\}$, and, for all $(w, f) \in W \times F$, $\pi(w, f) = 1$. Recall that at any stable outcome the workers (the firms) must obtain the same stable payoffs, i.e., stable payoffs are aligned. Consider the following unstable outcome (μ, u): worker w_1 and firm f_1 are matched, w_1 obtains a payoff $u_{w_1} = a \in [0, 1]$, f_1 obtains a payoff $u_{f_1} = 1 - a \in [0, 1]$, and the remaining agents w_2 and f_2 are single and obtain zero payoffs. Graphically, (μ, u) is represented as follows:

$$(\mu^1, u^1) \qquad \begin{array}{lcr} w_1 & \overset{a}{\longrightarrow}\overset{1-a}{\longleftarrow} & f_1 \\[2ex] w_2. & & .f_2 \end{array}$$

The set of agents that form couples is $C(\mu) = \{w_1, f_1\}$, w_1 and f_1 are optimal partners and receive stable payoffs, such that w_1 and f_1 are matched according to the stable outcome (μ^*, u^*) where

(i) $\mu^* = (w_1, f_1), (w_2, f_2)$ is an optimal matching, and

(ii) u^* is a stable payoff vector, i.e., $a \in (0, 1)$, [for all $i \in W$, $u_i^* = a$], and [for all $j \in F$, $u_j^* = 1 - a$].

Hence, following our blocking path (Step 3, Matching completion process) it suffices to match worker w_2 with firm f_2 with stable payoffs $u_{w_2} = a$ and $u_{f_2} = 1 - a$ in order to reach the stable outcome (μ^*, u^*) in one step. However, since payoffs are continuous, the probability that the blocking pair (w_2, f_2) is satisfied with exactly those stable payoffs is zero. Similarly as in Example 6 one can show that any path to stability would require such a "zero probability" alignment step. Hence, probabilistically, convergence to stability in a decentralized process cannot be obtained. △

Consider now the following notion of *ϵ-stability*. Let (W, F, π) be an assignment problem and u^* a stable payoff vector for (W, F, π). An outcome (μ, u) is ϵ-stable if μ is an optimal matching and all agents obtain payoffs in an ϵ-neighborhood of a stable payoff vector u^*, i.e., u is an ϵ-stable payoff vector if there exists a real number $\epsilon > 0$ such that, for all $i \in W \cup F$, $|u_i^* - u_i| < \epsilon$.

Suppose that along our blocking path an outcome (μ^{l+1}, u^{l+1}) is obtained from outcome (μ^l, u^l) by satisfying a blocking pair (w, f) with the condition that $u_w^{l+1} = u_w^*$ and $u_f^{l+1} = u_f^*$, i.e., payoffs must be aligned at the stable payoff vector u^*. We know that such a blocking pair will never be selected with positive probability in a decentralized process. However, if for $i \in \{w, f\}$ we consider the payoff $u_i^{l+1} \in (u_i^* - \epsilon, u_i^* + \epsilon)$ to be ϵ-stable, then a blocking pair (w, f) with payoffs $u_w^{l+1} \in (u_w^* - \epsilon, u_w^* + \epsilon)$ and $u_f^{l+1} \in (u_f^* - \epsilon, u_f^* + \epsilon)$ is selected with positive probability in a decentralized process. Hence, a decentralized process will converge to ϵ-stability with probability one. Furthermore, Theorem 1 induces a paths to ϵ-stability result without requiring Assumption 1.

Corollary 1 (ϵ-Stability). *Let (W, F, π) be an assignment problem and (μ, u) an arbitrary outcome for (W, F, π). Then, there exists a blocking path $(\mu^1, u^1), ..., (\mu^k, u^k)$ such that $(\mu, u) = (\mu^1, u^1)$ and (μ^k, u^k) is ϵ-stable. Furthermore, a randomly created path $(\mu^1, u^1), (\mu^2, u^2), ...$ converges with probability one to an ϵ-stable outcome.*

4.3 A Discussion of Three Closely Related Papers

Discretized Two-Sided Assignment with Weak Blocking: Chen et al. (2012) and Nax et al. (2013)

Our main result (Theorem 1) is related to Chen et al. (2012, Theorem 1) and Nax et al. (2013, Theorem 1) in that it implies the paths to stability results that they also obtain for their assignment model specifications.[8] Since both these papers obtain the same paths to stability result with its associated probabilistic interpretation (see our discussion in Section 4.2) using essentially the same proof technique, we explain the difference between their results and ours by referring to Chen et al. (2012).

Chen et al. (2012) study a labor market with finitely many heterogeneous workers and firms to illustrate the blocking dynamics in assignment problems. They prove the existence of blocking

[8]The existence of Chen et al.'s (2012) result was known to us in March 2011. Our results were publicly defended and published (Payot, 2011) in July 2011. We became aware of the result of Nax et al. (2013) only in 2012 (the earliest working paper version we saw is dated June 2012)

paths to stability for assignment problems, as we do. They make two main assumptions that make their model different from our model. First, Chen et al. (2012) consider an assignment problem with *discrete* side payments. Second, they use a *weak* blocking norm: two agents form a blocking pair if satisfying this blocking pair makes *at least* one of them strictly better off. As Chen et al. (2012) we also study an assignment problem with side payments. However, our model differs from theirs with respect to two dimensions. First, we consider an assignment problem with *continuous* side payments. Second, our *strict* blocking norm requires that two agents form a blocking pair if and only if satisfying this blocking pair makes both agents strictly better off. In contrast to Chen et al. (2012) we have shown that with our assumptions a blocking path to stability does not always exists in the assignment problem. We identified a necessary and sufficient condition to guarantee the existence of blocking paths to stability. As discussed in Section 4.2, the probabilistic interpretation of the path to stability result that Chen et al. (2012, Theorem 1) obtain does not apply to our continuous assignment model.

Although our results and the results of Chen et al. (2012) are closely related, we use a different proof technique. Chen et al. (2012) essentially adapt the proof strategy of Roth and Vande Vate (1990) to assignment problems with discrete side payments. They construct an algorithm that targets a side optimal outcome. Each time an agent, say agent i, is selected to block an outcome, he will choose to block with his most preferred partner, i.e., the blocking partner with whom he can generate the largest surplus, and offers this best blocking partner the smallest payoff consistent with the incentives to block, such that i obtains in the next outcome the largest payoff while forming a blocking pair. Hence, given that payoffs are polarized in the core, the payoffs of one side of the market monotonically increase whereas the payoffs of the other side monotonically decrease along the blocking path. Our proof is more akin to the one used by Diamantoudi et al. (2004) in the sense that our algorithm (as well as theirs) uses a target stable outcome to avoid cycling.

Despite the fact that our proof technique differs from the proof in Chen et al. (2012), our proof works well for their environment. To see how our blocking path construction works for assignment problems with discrete side payments is straight forward. If payoffs were discretized in our model, then any blocking surplus can only be divided in finitely many ways; switching from continuous to discrete payoffs in our model has the simplifying effect to reduce the number of blocking possibilities to a finite number. When applying our proof construction to a discretized assignment problem, our blocking path will still lead to stability but in possibly fewer steps (the reason why we then also can drop our necessary and sufficient assumption is that with discrete payoffs the convergence problems we indicate in Section 3 cannot occur). Hence, our result extends that of Chen et al. (2012) to assignment problems with continuous side payments.

Chen et al. (2012) allow weak blocking pairs to be formed. In contrast, we focused on strict payoff improvement for a blocking pair to be satisfied. Therefore, we need to target stable payoffs away from zero (and assume their existence by our necessary and sufficient Assumption 1) to ensure that agents that obtain zero payoffs at some point and are matched at a stable outcome have a clear incentive to form a blocking pair by obtaining a strictly higher payoff. This crucial situation of having to match zero payoff stable partners may appear multiple times along our blocking path.[9]

In our model, allowing for weak blocking pairs would simply make Assumption 1 unnecessary because then it would always be possible to satisfy a blocking pair of optimal partners with one of them being single even though the single agent obtains zero payoff at a stable outcome. For instance, recall that in Example 4 we constructed an infinite sequence of outcomes that converges to stability. The infiniteness of the sequence is precisely due to the fact that at a stable outcome the worker who is matched obtains a zero payoff. Thus, in this example a blocking path would exist under the assumption of weak blocking pairs.

Continuous One-Sided Assignment with Weak Blocking: Biró et al. (2013)

Biró et al. (2013) establish the existence of blocking paths to stability for one-sided assignment problems using the weak blocking norm. Hence, part of their model is more general than ours because they fully work out the one-sided assignment setting (we only remark that our proof technique can easily be used in a corresponding one-sided setting) and part of our model is more general since we use the more stringent strict blocking norm. Hence, our results are very close, but somewhat incomparable (in addition to being independently obtained).

Note that using a proof technique with a target stable outcome is absolutely necessary for one-sided assignment problems (for one-sided assignment problems the set of stable outcomes may be empty and if it is nonempty, then it does not need to form a complete lattice with two extreme points that reflect the polarization between both sides of the assignment problem). However, in spite of being a more general proof technique (since it also applies to one-sided assignment problems), the target outcome proof technique comes at the price of a less intuitive myopic blocking behavior along the constructed blocking path because at times blocking payoffs have to be aligned according to the stable target payoffs. In contrast, the classical two-sided proof technique used by Chen et al. (2012) and Nax et al. (2013) requires an actively blocking agent to find a blocking partner with the highest blocking surplus and then extract the highest possible blocking payoff from this blocking partner (with the weak blocking notion this corresponds to the active blocking partner taking the whole blocking surplus – this extreme blocking surplus extraction is not possible under the strict blocking norm that we use in contrast to the other articles we have been discussing in this section).

4.4 Median Stable Target Outcomes

In the previous section we have discussed that the classical "greedy" two-sided proof employed in Chen et al. (2012) and Nax et al. (2013) needs to be replaced by a target outcome proof in

[9]First, in Cases 2.2 and 3.2 within Step 2 (Stabilization Process) we have defined a short blocking sequence (2 steps) in order to rematch two optimal partners according to our stable target payoff vector. In the second phase of this short blocking sequence, we rematch one agent who forms a couple (but not with an optimal partner) with his optimal partner who is single and obtains zero payoff. Thus, without the weak blocking assumption, this rematching step is feasible only if a stable payoff away from zero exists (Assumption 1). Second, in Step 3 (Matching Completion Process) we complete the blocking path by matching single agents who are optimal partners. Given that all single agents obtain zero payoffs, Assumption 1 is needed in order to ensure a strict payoff improvement for each blocking agent.

our context because we use the strict blocking norm (another reason would be the consideration of the one-sided version of our model). Here we argue that this target method can be used in a centralized assignment market to stabilize an unstable outcome using a "compromise" target outcome. Recall that for two-sided assignment problems two extreme stable payoff vectors (and associated outcomes) exist: the worker-optimal stable payoff vector and the firm-optimal stable payoff vector. These are the stable payoff vectors that are most unequal within the set of stable payoff vectors. Schwarz and Yenmez (2011) define the median stable payoff vector (and associated outcomes) as a compromise solution and prove that they are well defined and exist.

Assume that in a centralized labor market we detect that current payoffs are not stable. A centralized adjustment process could then use the median stable payoff vector in our paths to stability algorithm to move to a stable outcome (and by doing so, the originally unstable payoffs, in the stabilization process, would be moved closer to the median stable payoffs). The rational behind such a centralized stabilization procedure would then be that the resulting outcome could be obtained via decentralized blocking that targets a compromise stable outcome for the current situation.

4.5 Concluding Remarks

We have studied two-sided one-to-one matching problems with continuous side payments. We have considered the existence of blocking paths to stability for such assignment problems under the strict blocking norm. In contrast to weak blocking paths results by Biró et al. (2013), Chen et al. (2012), and Nax et al. (2013), the existence of a blocking path to stability cannot always be guaranteed. We identified a necessary and sufficient condition (Assumption 1) for the existence of a blocking path to stability.

With Assumption 1, we distinguish between two types of stable outcomes for any given assignment problem: stable outcomes that involve matched agents with zero payoffs versus those stable outcomes where all matched agents receive strictly positive payoffs (recall that the role of zero here is that of an agent's reservation value in our normalized setup). We find that if stable outcomes are exclusively of the first type, then no path to stability exists (Theorem 2), while the existence of a stable outcome of the second type guarantees the existence of a path to stability (Theorem 1). Even when a path to stability is guaranteed to exist, our results show that finding or constructing such a path might not be trivial (the Proof of Theorem 1 demonstrates that the path construction could be rather involved and requires the use of a target stable outcome satisfying Assumption 1). Moreover, with examples such as Example 6 we show that an intuitively fair blocking dynamics might never converge to a stable outcome. These results seem to be bad news for stability as the result of our (myopic) decentralized process. However, we also show that if small deviations from stability are acceptable (i.e., ϵ-stability), then a randomly created path will converges with probability one to an ϵ-stable outcome (Corollary 1). Furthermore, our proof technique has the potential to be used in a centralized market in which a central planner may deliberately choose a specific stable target outcome, e.g., a median stable outcome, for the stabilization process described in the Proof of Theorem 1.

With suitable modifications of our model (i.e., allowing for weak blocking or specifying a discrete payoff structure), our results imply the results of Chen et al. (2012) and Nax et al.

(2013). However, the converse is not true: modifying their model to coincide with ours (i.e., imposing strict blocking and allowing continuous transfers) would not allow them to easily adapt their very different proofs to obtain our results. Our proof technique is somewhat similar to that of Biró et al. (2013). The main difference with Biró et al. (2013) is that we have to deal with the more stringent requirement of strict blocking, which is the reason why in our model a necessary and sufficient condition is added to obtain the existence of paths to stability. Even though we formulate our model as a two-sided model, our proof technique does not depend on the two-sidedness of the market and hence we could easily obtain corresponding results for a one-sided model *à la* Biró et al. (2013).

A Appendix: Proofs

Before we start the proof of Theorem 1 we introduce some notion concerning the reduction of matchings, payoff vectors, and outcomes. Let (W, F, π) be an assignment problem and (μ, u) an outcome for it. Recall that we denote the set of agents that form couples at matching μ by $C(\mu) := \{i \in W \cup F \mid \mu(i) \neq i\}$. Then, by $\mu|_{C(\mu)}$ we denote the *reduction of matching μ to the set of agents $C(\mu)$*; formally, $\mu|_{C(\mu)} \in \mathcal{M}(W \cap C(\mu), F \cap C(\mu))$ such that for all $i \in C(\mu)$, $\mu|_{C(\mu)}(i) := \mu(i)$. Similarly, by $u|_{C(\mu)}$ we denote the *reduction of payoff vector u to the set of agents $C(\mu)$*; formally, $u|_{C(\mu)} \in \mathbb{R}^{|C(\mu)|}$ such that $u|_{C(\mu)} := (u_i)_{i \in C(\mu)}$. Finally, $(\mu, u)|_{C(\mu)} = (\mu|_{C(\mu)}, u|_{C(\mu)})$ is the *reduction of outcome (μ, u) to the set of agents $C(\mu)$*. We say that the reduced outcome $(\mu, u)|_{C(\mu)}$ *is stable* if

(a) for all $i \in C(\mu)$, $u_i \geq 0$ and

(b) for all $(w, f) \in (W \cap C(\mu)) \times (F \cap C(\mu))$, $u_w + u_f \geq \pi(w, f)$.

Instead of saying that the reduced outcome $(\mu, u)|_{C(\mu)}$ is stable, we will also use the equivalent formulation that outcome (μ, u) is *stable within the set $C(\mu)$*.

Note that whenever we use the generic notation (i, j) for a pair, then $(i, j) \in W \times F$ or $i = j \in W \cup F$ are both possible. On a few occasions in the sequel we will also use the specific notation $\langle i, j \rangle$ when it is clear that either $(i, j) \in W \times F$ or $(j, i) \in W \times F$, but it is not important which is the case. With some abuse of notation we will not adjust the notation for the corresponding value $\pi(i, j)$ agents i and j create.

Proof of Theorem 1.
Let (W, F, π) be an assignment problem satisfying Assumption 1 and (μ, u) an arbitrary outcome for (W, F, π). By Assumption 1, there exists a stable outcome (μ^*, u^*) such that for each agent $i \in W \cup F$ who is not single, i.e., $\mu^*(i) \neq i$, we have

$$u_i^* > 0. \tag{*}$$

Step 1: Unmatch Process

We first unmatch as many couples as possible via blocking, i.e., we first maximize the number of single agents by matching blocking pairs (w, f) such that $\mu(w) \in F$ and $\mu(f) \in W$.[10] We construct the first part of our blocking path $(\mu, u) = (\mu^1, u^1), (\mu^2, u^2), \ldots$ as follows.

Step 1.l. For all $l \geq 1$, if there exists a blocking pair (w_l, f_l) for (μ^l, u^l) such that w_l and f_l are not single at μ^l, i.e., $w_l, f_l \in C(\mu^l)$,[11] then satisfy this blocking pair to obtain (μ^{l+1}, u^{l+1}). Note that the new set of agents that form couples $C(\mu^{l+1}) = C(\mu^l) \setminus \{\mu^l(w_l), \mu^l(f_l)\}$ contains fewer agents: $|C(\mu^{l+1})| = |C(\mu^l)| - 2$.

Since at each Step 1.l ($l \geq 1$) the number of agents that form couples is reduced by 2, the unmatch process reaches an outcome (μ^a, u^a) ($a \geq 1$) such that (μ^a, u^a) is stable within the set $C(\mu^a)$ in finitely many steps. The unmatch process (if starting from a non-empty matching) generates an outcome (μ^a, u^a) with at least one couple and $C(\mu^a) \neq \emptyset$.

After Step 1, we distinguish three cases for outcome (μ^a, u^a). The first one, Case (*), allows us to easily complete our blocking path using the stable target outcome (μ^*, u^*).

Case (*). There exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^a)} = (\mu^a, u^a)|_{C(\mu^a)}$ and for all $i \in C(\tilde{\mu}) \setminus C(\mu^a)$, $\tilde{u}_i > 0$. Then, set $(\mu^c, u^c) := (\mu^a, u^a)$ and go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to $(\tilde{\mu}, \tilde{u})$.

If outcome (μ^a, u^a) is not in Case (*), then we will apply Step 2 in order to appropriately stabilize the outcome through blocking. We distinguish two remaining cases for outcome (μ^a, u^a). First, Case (**) deals with boundary cases caused by agents who receive zero stable payoffs. Second, Case (***) deals with "classical" instability caused by non-optimal matching or by unstable payoffs.

Case (**). There exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^a)} = (\mu^a, u^a)|_{C(\mu^a)}$ and not Case (*). This implies that for some $j \in C(\tilde{\mu}) \setminus C(\mu^a)$, $\tilde{u}_j = 0$. If $C(\mu^a)$ contains only one couple, then go to Step 2 (Case 2.2). If $C(\mu^a)$ contains more than one couple, then go to Step 2 (Case 3.2). The reason why we might not be able to directly proceed with the completion of our blocking path by going to Step 3 is that it might require to match two singles such that one of them receives a zero payoff.[12] In the sequel, the payoff stabilization parts of Step 2 (Cases 2.2 and 3.2) will also deal with problematic zero payoffs.

Case (***). There does not exist any stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^a)} = (\mu^a, u^a)|_{C(\mu^a)}$. Then, the agents in $C(\mu^a)$ are not matched according to an optimal matching or they do not receive stable payoffs. We prove in Step 2 that then there exists a blocking path that "stabilizes" the set of agents that form couples, i.e., for the resulting outcome (μ^c, u^c)

[10] Had we modeled blocking paths to contain individually irrational outcomes and to allow singleton blockings, then we would in this step also unmatch any agent who receives a negative payoff at the initial outcome (μ, u).

[11] By (i) in the definition of an outcome, $\mu^l(w_l) \neq f_l$.

[12] An example of such a situation is: $W = \{w_1, w_2\}$, $F = \{f_1, f_2\}$, $\pi(w_i, f_j) = 1$ for all $i, j \in \{1, 2\}$, (μ^a, u^a) such that $C(\mu^a) = \{w_1, f_1\}$, and u^a such that $u^a_{w_1} = 1$ and $u^a_{w_2} = u^a_{f_1} = u^a_{f_2} = 0$. Proceeding as in the later Step 3 would require that w_2 and f_2 match with a zero payoff for f_2; this would not be a strict blocking pair.

there exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^c)} = (\mu^c, u^c)|_{C(\mu^c)}$ and such that we can then complete our blocking path to stability by matching single agents in $S(\mu^c)$ in Step 3.

The next step will use an induction argument to construct an outcome (μ^c, u^c) that belongs to Case (*).

Step 2: Stabilization Process

We continue our blocking path $(\mu^a, u^a), (\mu^{a+1}, u^{a+1}), ...$ with the aim to stabilize the set of agents that form couples. Throughout this step, we use a stable outcome (μ^*, u^*) satisfying inequality (*) for agents matched at μ^*. Note that whenever we refer to (μ^*, u^*) (and inequality (*)) we are applying Assumption 1.

We denote the number of couples at μ^a by $t = \frac{|C(\mu^a)|}{2}$ and consider the cases $t = 0$, $t = 1$, and $t > 1$. Note that strictly speaking, we could use the case $t = 0$ as our induction basis for the induction step $t > 0$ (instead of using $t = 1$ and $t > 1$). We explicitly add the case $t = 1$ for didactic reasons because some of the proof steps are more elementary and hence a good preparation for following the steps for $t > 1$. (Essentially, the following Case 2 could be omitted.)

Case 1 ($t = 0$). If $t = 0$, then $C(\mu^a) = \emptyset$ and $(\mu^a, u^a) = (\mu, u)$.[13] Hence, $W \cup F = S(\mu^a)$ and for all $i \in W \cup F$, $u_i^a = 0$, i.e., all agents are single and receive their reservation value. We set $(\mu^c, u^c) := (\mu^a, u^a)$ and go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to (μ^*, u^*) (using $(\tilde{\mu}, \tilde{u}) := (\mu^*, u^*)$ in Step 3).

Case 2 ($t = 1$). If $t = 1$, then $C(\mu^a) = \{w, f\}$ and $(w, f) \in \mu^a$ is the only couple at μ^a. Since there does not exist any stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^a)} = (\mu^a, u^a)|_{C(\mu^a)}$ as in Case (*) one of the following properties holds:

(2.1) worker w and firm f are not optimal partners, i.e., there does not exist an optimal matching $\tilde{\mu}$ such that $(w, f) \in \tilde{\mu}$ or

(2.2) either Case (**) or one of the agents $i \in \{w, f\}$ does not receive a stable payoff, i.e., $u_i^a \notin [\underline{u}_i^a, \overline{u}_i^a]$.

Case 2.1 (w and f are not optimal partners). Consider the optimal target matching μ^*. Consider the set $\{w, \mu^*(f), f, \mu^*(w)\}$ (in Case 3.1 we will denote the corresponding set by $C(\mu^a) \cup S^*(\mu^a)$). If $\mu^*(w) = w$ and $\mu^*(f) = f$, then, $\pi(w, f) = 0$, which would mean that changing μ^* by matching w with f would also yield an optimal matching; contradicting the fact that w and f are not optimal partners. Hence, $|\{w, \mu^*(f), f, \mu^*(w)\}| \in \{3, 4\}$.

Let $\bar{\mu}$ be the matching that is obtained from the optimal target matching μ^* by rematching agents in $\{w, \mu^*(f), f, \mu^*(w)\}$ according to matching μ^a, i.e.,

$$\bar{\mu}(i) = \begin{cases} \mu^a(i) & \text{if } i \in \{w, \mu^*(f), f, \mu^*(w)\} \\ \mu^*(i) & \text{otherwise.} \end{cases} \tag{2.1}$$

[13]Recall that the unmatch process generates an outcome (μ^a, u^a) with at least one couple.

Since w and f are not optimal partners, matching $\bar{\mu}$ is not optimal. Hence,

$$\sum_{(i,j)\in\bar{\mu}} \pi(i,j) < \sum_{(i,j)\in\mu^*} \pi(i,j). \tag{2.2}$$

First, assume that $|\{w, \mu^*(f), f, \mu^*(w)\}| = 4$, i.e., $\mu^*(w) \neq w$ and $\mu^*(f) \neq f$. Hence, $(w, f), (\mu^*(w), \mu^*(w)), (\mu^*(f), \mu^*(f)) \in \bar{\mu}$. By construction (2.1) of $\bar{\mu}$, matchings $\bar{\mu}$ and μ^* coincide for all agents $i \notin \{w, f, \mu^*(w), \mu^*_f\})$. Thus,

$$\underbrace{\pi(w,f)}_{u^a_w+u^a_f} + \underbrace{\pi(\mu^*(w),\mu^*(w))}_{=0=u^a_{\mu^*(w)}} + \underbrace{\pi(\mu^*(f),\mu^*(f))}_{=0=u^a_{\mu^*(f)}} < \pi(w,\mu^*(w)) + \pi(\mu^*(f),f). \tag{2.3}$$

Hence,

$$\left(u^a_w + u^a_{\mu^*(w)}\right) + \left(u^a_{\mu^*(f)} + u^a_f\right) < \pi(w,\mu^*(w)) + \pi(\mu^*(f),f). \tag{2.4}$$

Thus, $[u^a_w + u^a_{\mu^*(w)} < \pi(w, \mu^*(w))]$ or $[u^a_{\mu^*(f)} + u^a_f < \pi(\mu^*(f), f)]$. Then, $(w, \mu^*(w))$ or $(\mu^*(f), f)$ is a blocking pair of optimal μ^*-partners for outcome (μ^a, u^a).

Next, assume that $|\{w, \mu^*(f), f, \mu^*(w)\}| = 3$ such that $\mu^*(w) \neq w$ and $\mu^*(f) = f$. Thus, $(w, f), (\mu^*(w), \mu^*(w)) \in \bar{\mu}$ and by construction of $\bar{\mu}$ (matchings $\bar{\mu}$ and μ^* coincide for all agents $i \notin \{w, f, \mu^*(w)\}$),

$$\underbrace{\pi(w,f)}_{u^a_w+u^a_f} + \underbrace{\pi(\mu^*(w),\mu^*(w))}_{=0=u^a_{\mu^*(w)}} < \pi(w,\mu^*(w)) + \pi(f,f). \tag{2.3'}$$

Hence,

$$\left(u^a_w + u^a_{\mu^*(w)}\right) + \underbrace{u^a_f}_{\geq 0} < \pi(w,\mu^*(w)) + \underbrace{\pi(f,f)}_{=0}. \tag{2.4'}$$

Thus, $[u^a_w + u^a_{\mu^*(w)} < \pi(w, \mu^*(w))]$. Then, $(w, \mu^*(w))$ is a blocking pair of optimal μ^*-partners for outcome (μ^a, u^a).

Similarly as above, for $|\{w, \mu^*(f), f, \mu^*(w)\}| = 3$ such that $\mu^*(w) = w$ and $\mu^*(f) \neq f$ it follows that $(\mu^*(f), f)$ is a blocking pair of optimal μ^*-partners for outcome (μ^a, u^a).

To summarize, we can always identify a blocking pair $(w^*, f^*) \in \mu^*$ such that $(w^*, f^*) \in \{(w, \mu^*(w)), (\mu^*(f), f)\}$ for outcome (μ^a, u^a). Let (μ^{a+1}, u^{a+1}) be an outcome obtained by satisfying such a blocking pair (w^*, f^*) of optimal μ^*-partners $w^*, f^* \in C(\mu^{a+1})$. Note that $|C(\mu^{a+1})| = 2$.

If (as in Case (*)) there exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^{a+1})} = (\mu^{a+1}, u^{a+1})|_{C(\mu^{a+1})}$ and for all $i \in C(\tilde{\mu}) \setminus C(\mu^{a+1})$, $\tilde{u}_i > 0$, then set $(\mu^c, u^c) := (\mu^{a+1}, u^{a+1})$ and go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to $(\tilde{\mu}, \tilde{u})$. Otherwise (as in Cases (**) and (***)) set $(\mu^b, u^b) := (\mu^{a+1}, u^{a+1})$ and continue the blocking path as described next in Case 2.2.

Case 2.2 (optimal partners and either Case (**) or unstable payoffs). If at outcome (μ^a, u^a) agents w and f are optimal partners, then first set $(\mu^b, u^b) := (\mu^a, u^a)$. We continue the blocking path in this case with (μ^b, u^b) as the initial outcome (note that (μ^b, u^b) can come either directly from Step 1 or from Case 2.1 within Step 2). Most arguments for Case (**) and unstable payoff are the same [but whenever needed, we mark arguments specific to Case (**) in parentheses as illustrated here].

Agents w and f being optimal partners at μ^b implies that there exists an optimal matching $\tilde{\mu}^*$ such that $\tilde{\mu}^*(w) = f$ and $u^b_w + u^b_f = \pi(w, f)$. By Lemma 1 (b) and (μ^*, u^*) being a stable outcome, $(\tilde{\mu}^*, u^*)$ is also a stable outcome. Hence,

$$\text{for } (w, f) \in \mu^b, \; u^b_w + u^b_f = u^*_w + u^*_f = \pi(w, f). \tag{2.5}$$

Note that if $u^b_w = u^*_w$ and $u^b_f = u^*_f$, then we would be in Case (*) and not have reached Case 2.2. Hence, either $[u^b_w < u^*_w$ and $u^b_f > u^*_f]$ or $[u^b_w > u^*_w$ and $u^b_f < u^*_f]$.

Let $\bar{u}$ be the payoff vector that is obtained from the stable target payoff vector u^* by replacing the payoffs of worker w and firm f at u^* with those at u^b, i.e.,

$$\bar{u}_i = \begin{cases} u^b_i & \text{if } i \in \{w, f\} \\ u^*_i & \text{otherwise.} \end{cases} \tag{2.6}$$

[In Case (**), if payoff vector $\bar{u}$ is stable, then set $(\mu^c, u^c) := (\mu^b, u^b)$ and go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to $(\tilde{\mu}^*, \bar{u})$ (using $(\tilde{\mu}, \tilde{u}) := (\tilde{\mu}^*, \bar{u})$ in Step 3). Otherwise, there is at least one blocking pair $\langle i, j\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$.] Since agents w and f are not receiving stable payoffs at u^b, payoff vector $\bar{u}$ is also not stable and at least one blocking pair $\langle i, j\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$ exists. Since u^* is a stable payoff vector, by construction (2.6), any blocking pair for outcome $(\tilde{\mu}^*, \bar{u})$ involves either worker w or firm f.

Assume that agent $i \in \{w, f\}$ with $u^b_i > u^*_i$ is part of a blocking pair $\langle i, j\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$. Then, $\pi(i, j) > \bar{u}_i + \bar{u}_j = u^b_i + u^*_j > u^*_i + u^*_j$, contradicting the stability of payoff vector u^*. Hence, $u^b_i < u^*_i$ for the agent $i \in \{w, f\}$ who participates in blocking pair $\langle i, j\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$. Note that $j \in S(\mu^b)$ and $u^b_j = 0$. Thus, $\pi(i, j) > \bar{u}_i + \bar{u}_j \geq u^b_i + u^b_j$ and $\langle i, j\rangle$ is also a blocking pair for outcome (μ^b, u^b).

Satisfy this blocking pair to obtain the next outcome (μ^{b+1}, u^{b+1}) with the condition that $u^{b+1}_i \in (u^b_i, u^*_i)$. (Then, $u^{b+1}_j = \pi(i, j) - u^{b+1}_i > 0$.) Recall that $\{w, f\} = \{i, \mu^b(i)\}$, $\pi(i, \mu^b(i)) = u^*_i + u^*_{\mu^b(i)}$, and note that at outcome (μ^{b+1}, u^{b+1}) agent i's previous partner $\mu^b(i)$ is single and receives $u^{b+1}_{\mu^b(i)} = 0$. By construction, $u^{b+1}_i < u^*_i$ and $u^{b+1}_{\mu^b(i)} = 0 < u^*_{\mu^b(i)}$. Thus, $\langle i, \mu^b(i)\rangle = (w, f)$ is a blocking pair for outcome (μ^{b+1}, u^{b+1}) that we can satisfy to obtain outcome (μ^{b+2}, u^{b+2}) with the condition that $u^{b+2}_w = u^*_w$ and $u^{b+2}_f = u^*_f$.

At outcome (μ^{b+2}, u^{b+2}) agents w and f are optimal partners and they obtain stable payoffs that satisfy Case (*). Set $(\mu^c, u^c) := (\mu^{b+2}, u^{b+2})$ and go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to (μ^*, u^*) (using $(\tilde{\mu}, \tilde{u}) := (\mu^*, u^*)$ in Step 3).

Case 3 ($t > 1$). At outcome (μ^a, u^a), there are $t > 1$ couples and agents in $C(\mu^a)$ are not matched according to a stable outcome as in Case (*). We will use an induction argument to continue our blocking path $(\mu^a, u^a), (\mu^{a+1}, u^{a+1}), ...$ in order to construct an outcome (μ^c, u^c) that belongs to Case (*).

Induction Basis **($t = 1$).** For $t = 1$, we can construct a blocking sequence (μ^a, u^a), ..., (μ^c, u^c) (Case 2) such that the set of agents that form couples are stabilized as in Case (*), i.e., there exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^c)} = (\mu^c, u^c)|_{C(\mu^c)}$ and for all $i \in C(\tilde{\mu}) \setminus C(\mu^c)$, $\tilde{u}_i > 0$.

Induction Hypothesis **($t \geq 1$).** Assume that for $t \geq 1$, we can construct a blocking sequence $(\mu^a, u^a), ..., (\mu^c, u^c)$ such that the set of agents that form couples are stabilized as in Case (*), i.e., there exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^c)} = (\mu^c, u^c)|_{C(\mu^c)}$ and for all $i \in C(\tilde{\mu}) \setminus C(\mu^c)$, $\tilde{u}_i > 0$.

Induction Step **($t \to t+1$).** We now assume that at outcome (μ^a, u^a), there are $t+1 > 1$ couples and agents in $C(\mu^a)$ are not matched according to a stable outcome as in Case (*). Since there does not exist any stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^a)} = (\mu^a, u^a)|_{C(\mu^a)}$ and for all $i \in C(\tilde{\mu}) \setminus C(\mu^a)$, $\tilde{u}_i > 0$ one of the following holds:

(3.1) at least two agents that form a couple at μ^a are not optimal partners, i.e., for at least one couple $(w, f) \in \mu^a$ there does not exist an optimal matching $\tilde{\mu}$ such that $(w, f) \in \tilde{\mu}$ or

(3.2) either Case (**) or there exists an optimal matching $\tilde{\mu}$ such that $\tilde{\mu}|_{C(\mu^a)} = \mu^a|_{C(\mu^a)}$, but there does not exist a stable payoff vector $\tilde{u}$ such that $\tilde{u}|_{C(\mu^a)} = u^a|_{C(\mu^a)}$.

We now start to stabilize the set of agents that form couples as we did in Case 2. During this process, we might create blocking pairs that reduce the number of couples. Whenever this happens, we apply the induction hypothesis to obtain a blocking sequence that results in an outcome (μ^c, u^c) such that the set of agents that form couples are stabilized as in Case (*).

Case 3.1 (not all agents in $C(\mu^a)$ have optimal partners). Consider the optimal target matching μ^* and denote the set of single agents at μ^a that at μ^* are matched to agents in $C(\mu^a)$ by

$$S^*(\mu^a) := \{i \in S(\mu^a) \mid \mu^*(i) \in C(\mu^a)\}.$$

Note that each agent $i \in C(\mu^a) \cup S^*(\mu^a)$ has his optimal μ^*-partner in $C(\mu^a) \cup S^*(\mu^a)$.

Let $\bar{\mu}$ be the matching that is obtained from the optimal target matching μ^* by rematching agents in $C(\mu^a) \cup S^*(\mu^a)$ according to matching μ^a, i.e.,

$$\bar{\mu}(i) = \begin{cases} \mu^a(i) & \text{if } i \in C(\mu^a) \cup S^*(\mu^a) \\ \mu^*(i) & \text{otherwise.} \end{cases} \tag{3.1}$$

Since some agents that form couples at μ^a are not optimal partners, matching $\bar{\mu}$ is not optimal. Hence,

$$\sum_{(i,j)\in\bar{\mu}} \pi(i,j) < \sum_{(i,j)\in\mu^*} \pi(i,j). \tag{3.2}$$

By construction (3.1) of $\bar{\mu}$, matchings $\bar{\mu}$ and μ^* coincide for all agents $i \notin C(\mu^a) \cup S^*(\mu^a)$. Thus,

$$\underbrace{\sum_{\substack{(w,f)\in\mu^a \\ \text{s.t. } w,f\in C(\mu^a)}} \pi(w,f)}_{L1} + \underbrace{\sum_{\substack{(i,i)\in\mu^a \\ \text{s.t. } i\in S^*(\mu^a)}} \pi(i,i)}_{L2} < \underbrace{\sum_{\substack{(i,j)\in\mu^* \\ \text{s.t. } i,j\in C(\mu^a)\cup S^*(\mu^a)}} \pi(i,j)}_{R}. \tag{3.3}$$

Note that

$$\pi(i,j) = \begin{cases} u_w^a + u_f^a & \text{if } (i,j) = (w,f) \in \mu^a \text{ and } w,f \in C(\mu^a) \\ u_i^a = 0 & \text{if } (i,i) \in \mu^a \text{ and } i \in S^*(\mu^a). \end{cases}$$

This implies that we can rewrite $L1$ as

$$L1 = \sum_{\substack{(w,f)\in\mu^a \\ \text{s.t. } w,f\in C(\mu^a)}} \pi(w,f) = \sum_{\substack{(w,f)\in\mu^a \\ \text{s.t. } w,f\in C(\mu^a)}} (u_w^a + u_f^a) = \sum_{i\in C(\mu^a)} u_i^a$$

and $L2$ as

$$L2 = \sum_{\substack{(i,i)\in\mu^a \\ \text{s.t. } i\in S^*(\mu^a)}} \pi(i,i) = \sum_{\substack{(i,i)\in\mu^a \\ \text{s.t. } i\in S^*(\mu^a)}} u_i^a = \sum_{i\in S^*(\mu^a)} u_i^a.$$

Furthermore,

$$R = \sum_{\substack{(i,j)\in\mu^* \\ \text{s.t. } i,j\in C(\mu^a)\cup S^*(\mu^a)}} \pi(i,j) = \sum_{\substack{(w,f)\in\mu^* \\ \text{s.t. } w,f\in C(\mu^a)\cup S^*(\mu^a)}} \pi(w,f) + \sum_{\substack{(i,i)\in\mu^* \\ \text{s.t. } i\in C(\mu^a)\cup S^*(\mu^a)}} \pi(i,i).$$

We can now rewrite (3.3) as

$$\sum_{i\in C(\mu^a)\cup S^*(\mu^a)} u_i^a < \sum_{\substack{(w,f)\in\mu^* \\ \text{s.t. } w,f\in C(\mu^a)\cup S^*(\mu^a)}} \pi(w,f) + \sum_{\substack{(i,i)\in\mu^* \\ \text{s.t. } i\in C(\mu^a)\cup S^*(\mu^a)}} \pi(i,i). \tag{3.4}$$

Next, we map terms on the left side to terms on the right side of inequality (3.4) as follows (to be precise, we define a bijection between terms on the left and the right side of the inequality):

Note that for the terms associated with agents $i \in C(\mu^a) \cup S^*(\mu^a)$ such that $(i,i) \in \mu^*$ we have $u_i^a \geq 0 = \pi(i,i)$. Thus, in order for inequality (3.4) to hold, there must exist agents $w_1^*, f_1^* \in C(\mu^a) \cup S^*(\mu^a)$ such that $(w_1^*, f_1^*) \in \mu^*$ and

$$u_{w_1^*}^a + u_{f_1^*}^a < \pi(w_1^*, f_1^*).$$

Table 2: A bijection between terms on both sides of inequality (3.4).

term on the left side	for	term on the right side
$(u_w^a + u_f^a)$	$(w, f) \in \mu^*$ and $w, f \in C(\mu^a) \cup S^*(\mu^a)$	$\pi(w, f)$
u_i^a	$(i, i) \in \mu^*$ and $i \in C(\mu^a) \cup S^*(\mu^a)$	$\pi(i, i)$

Then, $(w_1^*, f_1^*) \in \mu^*$ is a blocking pair of optimal μ^*-partners for outcome (μ^a, u^a). By the definition of set $S^*(\mu^a)$ it follows that $w_1^* \in C(\mu^a)$ or $f_1^* \in C(\mu^a)$.

To summarize, we can always identify a blocking pair $(w_1^*, f_1^*) \in \mu^*$ such that $w_1^* \in C(\mu^a)$ or $f_1^* \in C(\mu^a)$ for outcome (μ^a, u^a). Let (μ^{a+1}, u^{a+1}) be the outcome obtained by satisfying such a blocking pair of optimal μ^*-partners $(w_1^*, f_1^*) \in C(\mu^{a+1})$. Note that $|C(\mu^{a+1})| \in \{|C(\mu^a)| - 2, |C(\mu^a)|\}$.

Next, if not all agents in $C(\mu^{a+1})$ have optimal partners, then we can repeat the same arguments to find another blocking pair $(w_2^*, f_2^*) \in \mu^*$ for outcome (μ^{a+1}, u^{a+1}) such that $w_2^* \in C(\mu^{a+1})$ or $f_2^* \in C(\mu^{a+1})$, etc., as follows:

Step 2.3.1.l. For all $l \geq 1$, if not all agents in $C(\mu^{a+l-1})$ have optimal partners, then let (μ^{a+l}, u^{a+l}) be the outcome obtained by satisfying a blocking pair $(w_l^*, f_l^*) \in \mu^*$ for outcome (μ^{a+l-1}, u^{a+l-1}) such that $w_l^* \in C(\mu^{a+l-1})$ or $f_l^* \in C(\mu^{a+l-1})$. Assume that $|C(\mu^{a+l})| = |C(\mu^{a+l-1})|$ and that (μ^{a+l}, u^{a+l}) is stable within the set $C(\mu^{a+l})$ (otherwise we apply the induction hypothesis and go to Step 3). Note that we have strictly increased the number of agents that form couples and are matched according to μ^*: $|C(\mu^{a+l}) \cap C(\mu^*)| = |C(\mu^{a+l-1}) \cap C(\mu^*)| + 2$.

Since at each Step 2.3.1.l ($l \geq 1$) the number of agents that form couples and are matched according to μ^* strictly increases by 2, we reach an outcome (μ^b, u^b) ($b > a$) where all agents in $C(\mu^b)$ have optimal partners in at most $t+1$ steps (unless we apply the induction hypothesis and go to Step 3).

If (as in Case (*)) there exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^b)} = (\mu^b, u^b)|_{C(\mu^b)}$ and for all $i \in C(\tilde{\mu}) \setminus C(\mu^b)$, $\tilde{u}_i > 0$, then set $(\mu^c, u^c) := (\mu^b, u^b)$ and go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to $(\tilde{\mu}, \tilde{u})$. Otherwise (as in Cases (**) and (***)), continue the blocking path as described in Case 3.2.

Case 3.2 (optimal partners and either Case (**) or unstable payoffs). If at outcome (μ^a, u^a) all couples are formed between optimal partners, then first set $(\mu^b, u^b) := (\mu^a, u^a)$. We continue the blocking path in this case with (μ^b, u^b) as the initial outcome (note that (μ^b, u^b) can come either directly from Step 1 or from Case 3.1 within Step 2). Most arguments for Case (**) and unstable payoff are the same [but whenever needed, we mark arguments specific to Case (**) in parentheses as illustrated here].

Each couple being formed by optimal partners at μ^b implies that there exists an optimal matching $\tilde{\mu}^*$ such that for all $i \in C(\mu^b)$, $\tilde{\mu}^*(i) = \mu^b(i)$ and for all $(w, f) \in \mu^b$, $u^b_w + u^b_f = \pi(w, f)$. By Lemma 1 (b) and (μ^*, u^*) being a stable outcome, $(\tilde{\mu}^*, u^*)$ is also a stable outcome. Hence,

$$\text{for all } (w, f) \in \mu^b, \; u^b_w + u^b_f = u^*_w + u^*_f = \pi(w, f). \tag{3.5}$$

Note that if for all $i \in C(\mu^b)$, $u^b_i = u^*_i$, then we would be in Case (*) and not have reached Case 3.2. Hence, for some $(w, f) \in \mu^b$, either $[u^b_w < u^*_w$ and $u^b_f > u^*_f]$ or $[u^b_w > u^*_w$ and $u^b_f < u^*_f]$.

Let $\bar{u}$ be the payoff vector that is obtained from the stable target payoff vector u^* by replacing the payoffs of the agents in $C(\mu^b)$ at u^* with those at u^b, i.e.,

$$\bar{u}_i = \begin{cases} u^b_i & \text{if } i \in C(\mu^b) \\ u^*_i & \text{otherwise.} \end{cases} \tag{3.6}$$

[In Case (**), if payoff vector $\bar{u}$ is stable, then set $(\mu^c, u^c) := (\mu^b, u^b)$ and go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to $(\tilde{\mu}^*, \bar{u})$ (using $(\tilde{\mu}, \tilde{u}) := (\tilde{\mu}^*, \bar{u})$ in Step 3). Otherwise, there is at least one blocking pair $\langle i, j\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$.] Since not all agents in $C(\mu^b)$ receive payoffs at u^b according to a stable payoff vector (recall that there does not exist a stable payoff vector $\tilde{u}$ such that $\tilde{u}|_{C(\mu^a)} = u^a|_{C(\mu^a)}$), payoff vector $\bar{u}$ is also not stable and at least one blocking pair $\langle i, j\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$ exists. Since u^* is a stable payoff vector, by construction (3.6), any blocking pair for outcome $(\tilde{\mu}^*, \bar{u})$ involves a matched agent $i_1 \in C(\mu^b)$ such that $u^b_{i_1} \neq u^*_{i_1}$.

Assume that agent $i_1 \in C(\mu^b)$ with $u^b_{i_1} > u^*_{i_1}$ is part of a blocking pair $\langle i_1, j_1\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$. Then, $\pi(i_1, j_1) > \bar{u}_{i_1} + \bar{u}_{j_1} = u^b_{i_1} + u^*_{j_1} > u^*_{i_1} + u^*_{j_1}$, contradicting the stability of payoff vector u^*. Hence, $u^b_{i_1} < u^*_{i_1}$ for the agent $i_1 \in C(\mu^b)$ who participates in blocking pair $\langle i_1, j_1\rangle$ for outcome $(\tilde{\mu}^*, \bar{u})$. Note that $j_1 \in S(\mu^b)$ and $u^b_{j_1} = 0$. Thus, $\pi(i_1, j_1) > \bar{u}_{i_1} + \bar{u}_{j_1} \geq u^b_{i_1} + u^b_{j_1}$ and $\langle i_1, j_1\rangle$ is also a blocking pair for outcome (μ^b, u^b).

Satisfy this blocking pair to obtain the next outcome (μ^{b+1}, u^{b+1}) with the condition that $u^{b+1}_{i_1} \in (u^b_{i_1}, u^*_{i_1})$. (Then, $u^{b+1}_{j_1} = \pi(i_1, j_1) - u^{b+1}_{i_1} > 0$.) Recall that $\pi(i_1, \mu^b(i_1)) = u^*_{i_1} + u^*_{\mu^b(i_1)}$ and note that at outcome (μ^{b+1}, u^{b+1}) agent i_1's previous partner $\mu^b(i_1)$ is single and receives $u^{b+1}_{\mu^b(i_1)} = 0$. By construction, $u^{b+1}_{i_1} < u^*_{i_1}$ and $u^{b+1}_{\mu^b(i_1)} = 0 < u^*_{\mu^b(i_1)}$. Thus, $\langle i_1, \mu^b(i_1)\rangle \in \mu^b$ is a blocking pair for outcome (μ^{b+1}, u^{b+1}) that we can satisfy to obtain outcome (μ^{b+2}, u^{b+2}) with the condition that $u^{b+2}_{i_1} = u^*_{i_1}$ and $u^{b+2}_{\mu^b(i_1)} = u^*_{\mu^b(i_1)}$.

To summarize, we can always identify two consecutive blocking pairs $\langle i_1, j_1\rangle$ (such that $i_1 \in C(\mu^b)$, $u^b_{i_1} < u^*_{i_1}$, and $j_1 \in S(\mu^b)$) and $\langle i_1, \mu^b(i_1)\rangle$ such that the resulting outcome rematches the original couple $\langle i_1, \mu^b(i_1)\rangle$ with payoffs $u^*_{i_1}$ and $u^*_{\mu^b(i_1)}$. After satisfying such a short stabilizing blocking sequence to obtain an outcome (μ^{b+2}, u^{b+2}), if either Case (**)

or there does not exist a stable payoff vector $\tilde{u}$ such that $\tilde{u}|_{C(\mu^{b+2})} = u^{b+2}|_{C(\mu^{b+2})}$, then we can repeat the same arguments to find another short stabilizing blocking sequence to obtain an outcome (μ^{b+4}, u^{b+4}), etc., as follows:

Step 2.3.2.l. For all $l \geq 1$, if either Case (**) or there does not exist a stable payoff vector $\tilde{u}$ such that $\tilde{u}|_{C(\mu^{b+2l-2})} = u^{b+2l-2}|_{C(\mu^{b+2l-2})}$, then let (μ^{b+2l}, u^{b+2l}) be the outcome obtained by satisfying a short stabilizing blocking sequence with blocking pairs $\langle i_l, j_l \rangle$ (such that $i_l \in C(\mu^{b+2l-2})$, $u^{b+2l-2}_{i_l} < u^*_{i_l}$, and $j_l \in S(\mu^{b+2l-2})$) and $\langle i_l, \mu^b(i_l) \rangle$ such that the resulting outcome rematches the original couple $\langle i_l, \mu^b(i_l) \rangle$ with payoffs $u^*_{i_l}$ and $u^*_{\mu^b(i_l)}$. Assume that $|C(\mu^{b+2l})| = |C(\mu^{b+2l-2})|$ (otherwise we apply the induction hypothesis and go to Step 3). Note that we have strictly increased the number of agents that form couples and receive payoffs according to u^*: $|\{i \in C(\mu^{b+2l}) \mid u^{b+2l}_i = u^*_i\}| = |\{i \in C(\mu^{b+2l-2}) \mid u^{b+2l-2}_i = u^*_i\}| + 2$.

Since at each Step 2.3.2.l ($l \geq 1$) the number of agents that form couples and receive payoffs according to u^* strictly increases by 2, we reach an outcome (μ^c, u^c) ($c > b$) in at most $t + 1$ steps (unless we apply the induction hypothesis and go to Step 3).

By construction, there now exists some stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^c)} = (\mu^c, u^c)|_{C(\mu^c)}$ and for all $i \in C(\mu^c)$, $u^c_i > 0$. We go to Step 3, where we complete our blocking path to stability by matching single agents in $S(\mu^c)$ according to $(\tilde{\mu}, \tilde{u})$.

Step 3: Matching Completion Process

We continue our blocking path $(\mu^c, u^c), (\mu^{c+1}, u^{c+1}), \ldots$ with the aim to rematch the single agents at μ^c according to the stable outcome $(\tilde{\mu}, \tilde{u})$ such that $(\tilde{\mu}, \tilde{u})|_{C(\mu^c)} = (\mu^c, u^c)|_{C(\mu^c)}$ and for all $i \in C(\tilde{\mu}) \setminus C(\mu^c)$, $\tilde{u}_i > 0$.

Step 3.l. For all $l \geq 1$, if there exists (w_l, f_l) such that $w_l, f_l \in C(\tilde{\mu}) \setminus C(\mu^{c+l-1})$ and $\tilde{\mu}(w^l) = f^l$, then (w_l, f_l) is a blocking pair for (μ^{c+l-1}, u^{c+l-1}) such that w_l and f_l are single at μ^{c+l-1}. Satisfy this blocking pair to obtain (μ^{c+l}, u^{c+l}) with the property that for $i \in \{w_l, f_l\}$, $u^{c+l}_i = \tilde{u}_i$. Note that the new set of agents that form couples $C(\mu^{c+l}) = C(\mu^{c+l-1}) \cup \{w_l, f_l\}$ contains more agents: $|C(\mu^{c+l})| = |C(\mu^{c+l-1})| + 2$.

The matching completion process increases the number of couples without perturbing the stability within the set of matched agents. The process terminates when all agents have been (re)matched according to the stable outcome $(\tilde{\mu}, \tilde{u})$. Hence, the matching completion process terminates in finitely many steps resulting in a stable outcome $(\tilde{\mu}, \tilde{u})$. □

Proof of Theorem 2.
Let (W, F, π) be an assignment problem violating Assumption 1. Hence, there exists no stable outcome (μ^*, u^*) such that for each agent $i \in W \cup F$ who is not single, i.e., $\mu^*(i) \neq i$, we have $u^*_i > 0$. Equivalently, for all stable outcomes (μ^*, u^*) there exists an agent $i \in W \cup F$ who is not single such that $u^*_i = 0$.

Define the set of agents who are not single but receive a zero payoff at some stable outcome by $X(W,F,\pi) = \{i \in W \cup F \mid \text{there exists } (\mu^*, u^*) \in \mathcal{S}(W,F,\pi) \text{ such that } \mu^*(i) \neq i \text{ and } u_i^* = 0\}$.

Let $(\hat{\mu}, \hat{u})$ be a stable outcome and let $(\tilde{\mu}, \tilde{u})$ be the outcome that is obtained by unmatching all agents $i \in X(W,F,\pi)$. Thus, $X(W,F,\pi) \subseteq S(\tilde{\mu})$. Then, from outcome $(\tilde{\mu}, \tilde{u})$ no blocking path leads to stability. The reason for this is that in order to end in a stable outcome (μ^*, u^*), one of the agents in $X(W,F,\pi)$ needs to be matched at a zero payoff along the blocking path. This, however, violates the strict blocking norm. □

B Appendix: Example for Theorem 1

Let (W,F,π) be an assignment problem given by $W = \{w_1, w_2, w_3, w_4\}$, $F = \{f_1, f_2, f_3, f_4\}$ and the characteristic function π is given (in matrix notation) by

$$\pi = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$$

where the element in row i and column j corresponds to the value $\pi(w_i, f_j)$.[14] Let (μ^*, u^*) be a stable outcome for (W,F,π). The unique optimal matching for (W,F,π) is

$$\mu^* = [(w_1, f_1), (w_2, f_2), (w_3, f_3), (w_4, f_4)]$$

and the set of stable payoffs is such that

(i) for all $(w,f) \in \mu^*$, $u_w^* + u_f^* = 2$ and

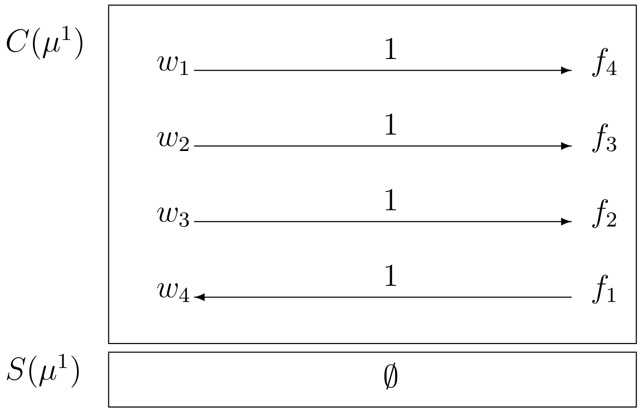
(ii) for all $(w,f) \in W \times F$, $u_w^* + u_f^* > 1$.

Initial Outcome: consider the unstable outcome (μ^1, u^1), such that

$$\mu^1 = [(w_1, f_4), (w_2, f_3), (w_3, f_2), (w_4, f_1)] \text{ and}$$
$$u^1 = (u_{w_1}^1, u_{w_2}^1, u_{w_3}^1, u_{w_4}^1, u_{f_1}^1, u_{f_2}^1, u_{f_3}^1, u_{f_4}^1) = (1,1,1,0,1,0,0,0).$$

Thus, the sets of matched agents and single agents at μ^1 are $C(\mu^1) = W \cup F$ and $S(\mu^1) = \emptyset$, respectively. The picture below represents outcome (μ^1, u^1).

[14]For instance, worker w_2 and firm f_2 generate value $\pi(w_2, f_2) = 2$ and worker w_2 and firm f_3 generate value $\pi(w_2, f_3) = 1$.



There are five blocking pairs for (μ^1, u^1), all of them being within $C(\mu^1)$:

$$\begin{aligned}
&(w_2, f_2),\\
&(w_3, f_3),\\
&(w_4, f_2),\ (w_4, f_3),\ \text{and } (w_4, f_4).
\end{aligned}$$

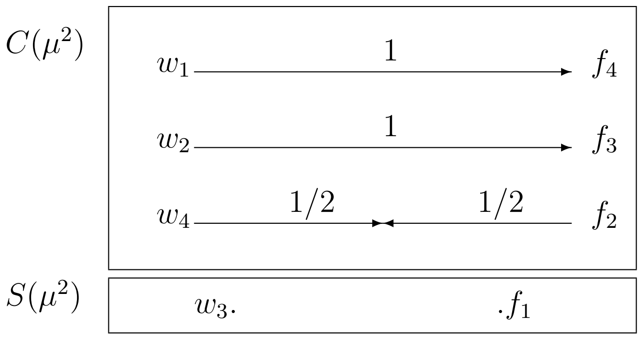
Step 1: Unmatch Process

First we unmatch as many couples as possible.

Step 1.1. Satisfy the blocking pair (w_4, f_2) to obtain outcome (μ^2, u^2), such that w_4 and f_2 equally split the blocking surplus

$$bs(\mu^1, u^1; w_4, f_2) = \pi(w_4, f_2) - u^1_{w_4} - u^1_{f_2} = 1.$$

Hence, at outcome (μ^2, u^2), w_4 and f_2 form a couple and obtain payoffs $u^2_{w_4} = 1/2$ and $u^2_{f_2} = 1/2$, and their previous partners w_3 and f_1 are single and obtain zero payoffs. All the other agents are matched to the same partners and earn the same payoffs as before. Outcome (μ^2, u^2) is thus



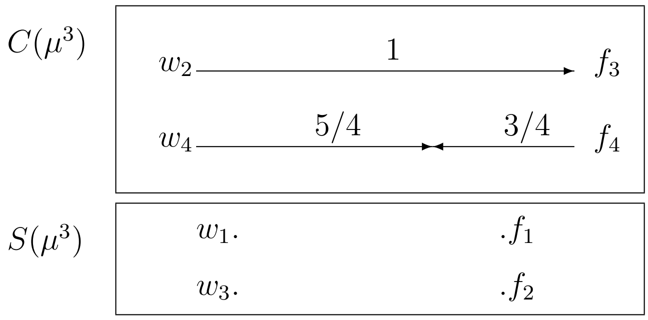
There are three blocking pairs for (μ^2, u^2) within $C(\mu^2)$:

$$\begin{aligned}
&(w_2, f_2),\\
&(w_4, f_3),\ \text{and } (w_4, f_4).
\end{aligned}$$

Step 1.2. Satisfy the blocking pair (w_4, f_4) to obtain outcome (μ^3, u^3), such that w_4 and f_4 equally split the blocking surplus

$$bs(\mu^2, u^2; w_4, f_4) = \pi(w_4, f_4) - u^2_{w_4} - u^2_{f_4} = 3/2.$$

Hence, at outcome (μ^3, u^3), w_4 and f_4 form a couple and obtain payoffs $u^3_{w_4} = 5/4$ and $u^3_{f_4} = 3/4$, and their previous partners w_1 and f_2 are single and obtain zero payoffs. All the other agents are matched to the same partners and earn the same payoffs as before. Outcome (μ^3, u^3) is thus



Outcome (μ^3, u^3) is stable within the set of agents that form couples, i.e., $(\mu^3, u^3)|_{C(\mu^3)}$ is stable. Then, set $(\mu^a, u^a) := (\mu^3, u^3)$ and go to Step 2 where we stabilize the set of agents that form couples.

Step 2: Stabilization Process

Second, we stabilize the set of agents that form couples. For instance, two blocking pairs of optimal partners for (μ^a, u^a): $(w_2, \mu^*(w_2)) = (w_2, f_2)$ and $(w_3, \mu^*(w_3)) = (w_3, f_3)$. If we satisfy one of those blocking pairs, we increase the number of agents who are matched to an optimal partner. As we will see, depending on which blocking pair we satisfy, that is either (w_2, f_2) or (w_3, f_3), the blocking path might take different routes. We investigate the two cases.

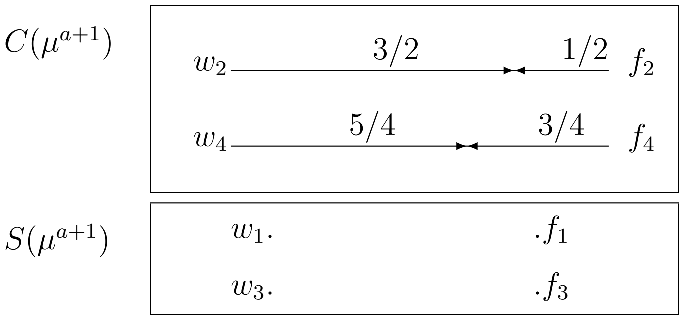
Case 1: Let (w_2, f_2) block (μ^a, u^a)

Satisfy the blocking pair (w_2, f_2) to obtain outcome (μ^{a+1}, u^{a+1}), such that w_2 and f_2 equally split the blocking surplus

$$bs(\mu^a, u^a; w_2, f_2) = \pi(w_2, f_2) - u^a_{w_2} - u^a_{f_2} = 1.$$

Hence, at outcome (μ^{a+1}, u^{a+1}), w_2 and f_2 form a couple and obtain payoffs $u^{a+1}_{w_2} = 3/2$ and $u^{a+1}_{f_2} = 1/2$, and w_2's previous partner f_3 is single and obtains zero payoffs. All the other agents are matched to the same partners and earn the same payoffs as before.

Outcome (μ^{a+1}, u^{a+1}) is thus



Outcome (μ^{a+1}, u^{a+1}) is stable within the set of agents that form couples, i.e., $(\mu^{a+1}, u^{a+1})|_{C(\mu^{a+1})}$ is stable. Furthermore, all agents in $C(\mu^{a+1})$ are matched to an optimal partner. Let

$$\tilde{u} = (\frac{3}{2}, \frac{3}{2}, \frac{3}{2}, \frac{5}{4}, \frac{3}{4}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$$

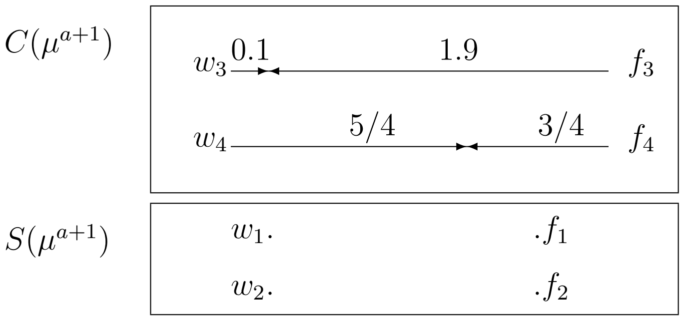
be a payoff vector. Notice that $\tilde{u}$ is stable since, for all $(w, f) \in W \times F$, $\tilde{u}_w + \tilde{u}_f \geq \pi(w, f)$. Hence, by Lemma 1, outcome $(\mu^*, \tilde{u})$ is stable. Furthermore, we are in Case (*) with optimal blocking pairs

$$(w_1, f_1) \text{ and } (w_3, f_3)$$

for (μ^{a+1}, u^{a+1}) within $S(\mu^{a+1})$. Set $(\mu^c, u^c) := (\mu^{a+1}, u^{a+1})$ and go to Step 3.

Case 2: Let (w_3, f_3) **block** (μ^a, u^a)

Satisfy the blocking pair (w_3, f_3) to obtain outcome (μ^{a+1}, u^{a+1}), such that w_3 and f_3 obtain payoffs $u^{a+1}_{w_3} = 0.1$ and $u^{a+1}_{f_3} = 1.9$. Hence, at outcome (μ^{a+1}, u^{a+1}), w_3 and f_3 form a couple, and f_3's previous partner w_2 is single and obtains zero payoffs. All the other agents are matched to the same partners and earn the same payoffs as before. Outcome (μ^{a+1}, u^{a+1}) is thus



Outcome (μ^{a+1}, u^{a+1}) is not stable within the set of agents that form couples, i.e., $(\mu^{a+1}, u^{a+1})|_{C(\mu^{a+1})}$ is not stable. Precisely, worker w_3 and firm f_4 generate a positive blocking surplus

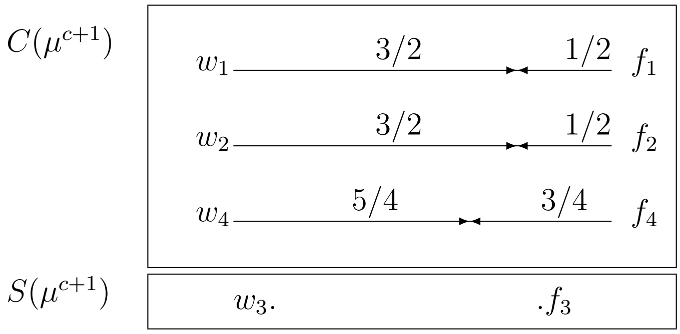
$$bs(\mu^{a+1}, u^{a+1}; w_3, f_4) = \pi(w_3, f_4) - u^{a+1}_{w_3} - u^{a+1}_{f_4} = 1 - 0.1 - \frac{3}{4} = 0.15,$$

such that (w_3, f_4) is a blocking pair for (μ^{a+1}, u^{a+1}). Then, satisfy this blocking pair to reduce the set of agents that form couples by one couple and apply the *induction hypothesis* to obtain a blocking sequence that stabilizes the set of agents that form couples.

Step 3: Matching Completion Process

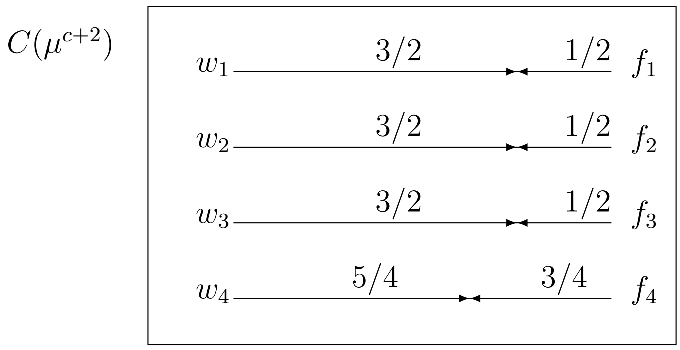
Let $(\mu^c, u^c) := (\mu^{a+1}, u^{a+1})$ (obtained from Case 1). In this step, we complete the blocking path to stability by matching the single agents in $S(\mu^c)$ according to the stable outcome $(\mu^*, \tilde{u})$.

Step 3.1. Satisfy the blocking pair (w_1, f_1) to obtain outcome (μ^{c+1}, u^{c+1}), such that w_1 and f_1 obtain stable payoffs $u^{c+1}_{w_1} = \tilde{u}_{w_1} = 3/2$ and $u^{c+1}_{f_1} = \tilde{u}_{f_1} = 1/2$. Hence, at outcome (μ^{c+1}, u^{c+1}), w_1 and f_1 form a couple and obtain stable payoffs, and the remaining single agents are w_3 and f_3 who obtain zero payoffs. All the other agents are matched to the same partners and earn the same payoffs as before. Outcome (μ^{c+1}, u^{c+1}) is thus



Outcome (μ^{c+1}, u^{c+1}) is stable within the set of agents that form couples, i.e., $(\mu^{c+1}, u^{c+1})|_{C(\mu^{c+1})}$ is stable. Furthermore, all agents that form couples are matched to an optimal partner and obtain stable payoffs. Thus, there is only one blocking pair for (μ^{c+1}, u^{c+1}): (w_3, f_3).

Step 3.2. Satisfy the blocking pair (w_3, f_3) to obtain outcome (μ^{c+2}, u^{c+2}), such that w_3 and f_3 obtain stable payoffs $u^{c+2}_{w_3} = \tilde{u}_{w_3} = 3/2$ and $u^{c+2}_{f_3} = \tilde{u}_{f_3} = 1/2$. Hence, at outcome (μ^{c+2}, u^{c+2}), w_3 and f_3 form a couple and obtain stable payoffs, and there are no single agents left. All the other agents are matched to the same partners and earn the same payoffs as before. Outcome (μ^{c+2}, u^{c+2}) is thus



$S(\mu^{c+2})$	$\emptyset$

At outcome (μ^{c+2}, u^{c+2}), all the agents are matched to an optimal partner and obtain stable payoffs. Therefore, outcome (μ^{c+2}, u^{c+2}) is stable and the blocking path $(\mu^1, u^1), ..., (\mu^{c+2}, u^{c+2})$ leads to stability in finitely many steps.

References

Abdulkadiroğlu, A., Pathak, P. A., and Roth, A. E. (2005): "The New York City high school match." *The American Economic Review*, 95(2): 364–367.

Abdulkadiroğlu, A. and Sönmez, T. (2003): "School choice: A mechanism design approach." *The American Economic Review*, 93(3): 729–747.

Ausubel, L. M. (2006): "An efficient dynamic auction for heterogeneous commodities." *The American Economic Review*, 96(3): 602–629.

Biró, P., Bomhoff, M., Golovach, P. A., Kern, W., and Paulusma, D. (2013): "Solutions for the stable roommates problem with payments." In *Graph-Theoretic Concepts in Computer Science*, Lecture Notes in Computer Science, pages 69 –80. Springer Berlin Heidelberg.

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

Crawford, V. P. (2008): "The flexible-salary match: a proposal to increase the salary flexibility of the national resident matching program." *Journal of Economic Behavior and Organization*, 66(2): 149–160.

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

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

Demange, G., Gale, D., and Sotomayor, M. (1986): "Multi-item auctions." *The Journal of Political Economy*, 94(4): 863–872.

Diamantoudi, E., Miyagawa, E., and Xue, L. (2004): "Random paths to stability in the roommate problem." *Games and Economic Behavior*, 48(1): 18–28.

Gale, D. and Shapley, L. S. (1962): "College admissions and the stability of marriage." *The American Mathematical Monthly*, 69(1): 9–15.

Gul, F. and Stacchetti, E. (2000): "The English auction with differentiated commodities." *Journal of Economic Theory*, 92(1): 66–95.

Hatfield, J. W., Kominers, S. D., Nichifor, A., Ostrovsky, M., and Westkamp, A. (2013): "Stability and Competitive Equilibrium in Trading Networks." *Journal of Political Economy (forthcoming).*

Kelso, J., Alexander S and Crawford, V. P. (1982): "Job matching, coalition formation, and gross substitutes." *Econometrica*, 50(6): 1483–1504.

Klaus, B. and Klijn, F. (2007): "Paths to stability for matching markets with couples." *Games and Economic Behavior*, 58(1): 154–171.

Knuth, D. E. (1976): *Mariages stables.* Les Presses de l'Université de Montréal.

Milgrom, P. (2000): "Putting auction theory to work: the simultaneous ascending auction." *The Journal of Political Economy*, 108(2): 245–272.

Nax, H. H., Pradelsky, B. S. R., and Young, H. P. (2013): "The evolution of core stability in decentralized matching markets." Working paper, University of Oxford.

Payot, F. (2011): *Three essays in economics of innovation and matching theory.* Ph.D. thesis, University of Lausanne. Chapter 3 joint with Bettina Klaus, Public Thesis Defense July 14th.

Roth, A. E. (1984): "The evolution of the labor market for medical interns and residents: a case study in game theory." *The Journal of Political Economy*, 92(6): 991–1016.

Roth, A. E. and Sotomayor, M. A. O. (1990): *Two-sided matching: A study in game-theoretic modeling and analysis.* Cambridge University Press.

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

Schwarz, M. and Yenmez, B. (2011): "Median stable matchings for markets with wages." *Journal of Economic Theory*, 146: 619–637.

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

Sotomayor, M. (2003): "Some further remark on the core structure of the assignment game." *Mathematical Social Sciences*, 46(3): 261–265.

Sun, N. and Yang, Z. (2009): "A double-track adjustment process for discrete markets with substitutes and complements." *Econometrica*, 77(3): 933–952.

Wako, J. (2006): "Another proof that assignment games have singleton cores only if multiple optimal matchings exist." *Economic Theory*, 29(1): 213–217.