CrossMark

Transfers and exchange-stability in two-sided matching problems

Emiliya Lazarova[1] · **Peter Borm[2]** ·
Arantza Estévez-Fernández[3]

Abstract In this paper we consider one-to-many matching problems where the preferences of the agents involved are represented by monetary reward functions. We characterize Pareto optimal matchings by means of contractual exchange stability and matchings of maximum total reward by means of compensational exchange stability. To conclude, we show that in going from an initial matching to a matching of maximum total reward, one can always provide a compensation schedule that will be ex-post stable in the sense that there will be no subset of agents who can all by deviation obtain a higher reward. The proof of this result uses the fact that the core of an associated compensation matching game with constraints is nonempty.

Keywords Matching · Pareto optimal matching · Contractual exchange stability · Compensational stability · Compensation schedule

JEL Classification C78 · C71 · D60

✉ Emiliya Lazarova
E.Lazarova@uea.ac.uk

Peter Borm
P.E.M.Borm@uvt.nl

Arantza Estévez-Fernández
arantza.estevezfernandez@vu.nl

[1] School of Economics, University of East Anglia, Norwich, UK

[2] Department of Econometrics and Operations Research, Tilburg University, Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

[3] Tinbergen Institute and Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

🙆 Springer

1 Introduction

Restructuring is an ongoing process in today's dynamic business world where companies are constantly looking for ways to enhance efficiency, to gain competitive edge, and to keep up with industry trends. A tension exists, however, between a job allocation that is best from an individual worker's or department's point of view and that of the organization as a whole. Thus, centralized restructuring that involves moving staff across different departments in an institution, potentially, leaves some staff worse-off. When a restructuring schedule is devised one should consider, therefore, the opportunity for an ex-post reorganization by some departments or workers that may undermine the process as a whole.

The focus of study is the class of Pareto optimal matchings of workers to departments within an organization, and, within this set, we pay particular attention to those matchings that maximize an organization's overall reward. We characterize the set of Pareto optimal matchings and the set of matchings that maximize the institution's total reward by means of two stability notions: contractual exchange stability and compensational exchange stability, respectively.

Contractual exchange stability requires that any deviation, i.e. workers changing their assigned department or departments exchanging subsets of workers, is approved by all affected parties. Underlying this stability notion is the assumption that an agent grants approval to a deviation only if his reward does not decrease as a result of the change. Our notion of contractual exchange stability is in the spirit of the notion of contractual individual stability introduced by Bogomolnaia and Jackson (2002) and analyzed in the context of hedonic coalition formation games. Like contractual exchange stability, the notion of individual contractual stability presumes that an agent can leave her current coalition only if her group mates in that coalition agree to the move. These agents on their part agree only if they are at least as well off without the leaving member of the coalition as when she is in. Two aspects distinguish our study of contractual stability from that by Bogomolnaia and Jackson (2002): the cardinal representation of preferences and the deviation possibilities. The former difference is inconsequential as only the ordinal properties of agents' preferences are utilized in the analysis of contractual exchange stability. The latter difference is driven by the fact that we focus on a two-sided many-to-one matching environment which is intrinsically more restrictive than that of coalition formation: for example, same-sided agents cannot be matched to each other without also being matched to an agent from the opposite side in a matching problem whereas any subset of agents can form a coalition in a coalition formation problem.

The cardinal representation of agents' preferences is instead essential in the definition of compensation exchange stability. Compensation exchange stability allows the deviating parties to compensate the agents that are made worse-off by the change. Thus, a deviation is viable only if the net increase in rewards of the deviating agents exceeds the total losses of those agents whose rewards decrease due to the deviation. Clearly, some deviations which are not allowed under the notion of contractual exchange stability (because they leave some agents worse-off) are possible under the notion of compensational exchange stability (because those agents can be compensated).

Our notion of compensation exchange stability, just like the notion of contractual exchange stability, has its counterpart in the coalition formation literature. For example, Lazarova et al. (2011) use a similar notion in the context of coalitional games. These authors, however, focus on unilateral deviations in that context, whereas here deviations may involve multiple agents. Like in hedonic coalition formation games discussed above, in coalitional games there are no restrictions on the set of coalitions that may form, whereas in a matching problem such restrictions exist.

An important class of two-sided matching problems with a cardinal representation of agents' preferences is the assignment problem. Seminal works that discuss algorithms for finding assignments with maximum total rewards are Knuth (1955) and Koopmans and Beckmann (1957) while Shapley and Shubik (1971) seminally define and study the notion of an assignment game.[1] In an assignment problem every *tuple* of matched agents is characterized by a cardinal value that is taken to represent the productivity or overall surplus that these agents can generate when matched together. In contrast, in the matching environment studied here every *agent* in a matched tuple is assigned a cardinal reward that represents the (share of the) value she is guaranteed in the matching. Our choice of assumptions is motivated, on the one hand, by our research question—the characterization of the set of Pareto optimal matchings and the set of matchings that maximize the agents' total reward, respectively by means of stability notions—as it allows us to abstract away from the problem of how the joint surplus is shared among the matched agents and focus on the specificities of the deviation possibilities instead. On the other hand, it is motivated by the particular application that we have in mind—institutional organization and restructuring—where workers are guaranteed a salary and departments face well-defined budgets.

This setting also directs our search for appropriate stability notions to be used in the analysis. Our notions of stability are related to the notion of exchange stability developed by Alcalde (1995) in the context of one-to-one matching problems. Gale and Shapley (1962) establish the concept of pairwise stability, which is adopted by Shapley and Shubik (1971) in the analysis of assignment games and the posterior literature. An important difference between the notions of exchange stability and those used in the matching literature is that in the former agents are only allowed to "swap" partners (i.e. the deviation involves agents from the same side) and are not allowed to opt out (i.e., unmatched agents are not considered) whereas in the latter agents are only allowed to "divorce", i.e. opt out of a matching to be unmatched or deviate to a matching with a new partner from the opposite side. Importantly, Alcalde (1995) shows the independence of the set of exchange stable outcomes and the set of pairwise stable outcomes in the context of a standard one-to-one matching problem where no transfers between agents are allowed. Like Alcalde (1995), we take the view that an analysis based on exchange stability should be taken as complementary to that based on pairwise stability and that the choice of a stability notion should be driven by the research focus. In this study, our choice of stability notions is driven by our task to

---

[1] We refer the reader to Pentico (2007) for a recent survey and classification of the literature on the assignment problem.

characterize two desirable sets of outcomes from an institution's point of view.[2] Last, we point out that our framework for exchange stability is different from that of Alcalde (1995) in the sense that we consider many-to-one matching problems with peer effects (i.e. the reward of a worker in a matching depends on the group of coworkers assigned to the same department) and, moreover, allow for monetary rewards. In the former aspect our work is related to the literature on many-to-one matching problems with peer effects (cf. Dutta and Massó (1997) and more recently Pycia (2012)). We differ from these works in the adopted stability notions (in our case this is exchange stability, whereas they usually consider extensions of pairwise stability) and again in our focus on the sets of matching that maximize the total agents' reward (as the many-to-one matching literature assumes ordinal preferences). Among the generalized assignment games works, e.g. Sotomayor (1999), on the other hand, we are not aware of any study where peer effects are modelled.

The adopted framework of cardinal representation of preferences allows us to also address the question if, given an initial matching of workers to departments, there exists a compensation schedule such that a centralized restructuring towards a matching of maximum total reward of the organization will be ex-post stable, in the sense that there will be no subset of workers or departments who can all by deviation obtain a higher reward. We analyze this question by introducing a transferable utility cooperative game, called a compensation matching game, and by showing that each compensation matching game has a nonempty core, which ensures that a compensation schedule as described above exists.

The paper is structured as follows. Section 2 provides notions on matchings for matching problems with monetary rewards and peer effects used throughout the paper. In Sect. 3, we characterize Pareto optimality of matchings by means of contractual exchange stability on the class of strict matching problems. In Sect. 4, we characterize matchings with maximum total reward by means of compensational exchange stability. Section 5 analyzes the existence of stable compensation schedules when an organization restructures the initial situation towards a matching of maximum total reward.

2 Basic notions

We consider two distinct types of economic agents: workers and departments. For ease of notation, we use lower case letters to denote a generic member i within the set of workers $\mathcal{N}$, and upper case letters to denote a generic department H within the set of departments $\mathcal{D}$. A vector of natural numbers $q = (q_H)_{H \in \mathcal{D}}$ represents the capacity of each department with q_H being the capacity of department H with $q_H \geq 1$. We assume that $|\mathcal{N}| = \sum_{H \in \mathcal{D}} q_H$, i.e. the total number of workers equals the total capacity available at the departments. An allocation of workers to departments forms a partition of the set of workers in $\mathcal{N}$, $P = \{P_H\}_{H \in \mathcal{D}}$, such that the size of the partition

elements is consistent with the departmental capacity, i.e. $|P_H| = q_H$ for all $H \in \mathcal{D}$. The set of all such partitions is denoted by $\mathcal{P}(\mathcal{N}, \mathcal{D}, q)$. If no confusion arises, we write $\mathcal{P}$ instead of $\mathcal{P}(\mathcal{N}, \mathcal{D}, q)$.

An allocation of workers to departments is given by a *matching* μ where $\mu : \mathcal{N} \to \mathcal{D}$ such that $P(\mu) = \{\mu^{-1}(H)\}_{H \in \mathcal{D}} \in \mathcal{P}$. The class of matchings is denoted by $\mathcal{M}(\mathcal{N}, \mathcal{D}, q)$. If no confusion arises, we write $\mathcal{M}$ instead of $\mathcal{M}(\mathcal{N}, \mathcal{D}, q)$. Notice that in our definition of a matching, we do not allow for workers to remain unmatched and for departments to have unfilled capacity. The choice of our assumption is driven by the particular environment that we have in mind, where existing workers are allocated to existing departments and firing or a closure of a department are not a possibility. Alcalde (1995) study a model with similar characteristics. This assumption is also common in the assignment problem literature, where the goal is to find an assignment with the highest total value (see Pentico 2007 for a review). As our focus is on the set of Pareto optimal matchings and the set of matchings of maximum total reward (formally defined below), the assumption is inconsequential as long as any outcome in which a department has unfilled capacity or a worker is unmatched yields strictly lower reward than any other outcome where all departments have full capacity and all workers are matched to a department.[3]

We assume that for each worker a matching leads to a monetary reward which depends on the department to which he is assigned and the identity of his peers in that department. Thus, for each $i \in \mathcal{N}$ there is a monetary reward function r_i where for each $(\mathcal{S}_i, H)$ with $H \in D$ and $\mathcal{S}_i \in 2^{N \setminus \{i\}}$ such that $|\mathcal{S}_i| = q_H - 1$, $r_i(\mathcal{S}_i, H)$ can be read as the reward that worker i receives when matched to department H with a set of coworkers $\mathcal{S}_i \subseteq \mathcal{N} \setminus \{i\}$.

Similarly, we assume that for each department $H \in \mathcal{D}$, there is a monetary reward function $\pi_H : \{\mathcal{S} \subset \mathcal{N} \mid |\mathcal{S}| = q_H\} \to \mathbb{R}$.

Given a matching $\mu \in \mathcal{M}$, we will use the shorthand notation $r_i(\mu)$ to denote the monetary reward obtained by worker $i \in \mathcal{N}$ when the department to which i belongs and the set of i's coworkers are given by matching μ, i.e. $r_i(\mu) = r_i(\mu^{-1}(\mu(i)) \setminus \{i\}, \mu(i))$. Similarly, $\pi_H(\mu)$ denotes the reward obtained by department $H \in \mathcal{D}$ from the set of workers assigned by μ to H, i.e. $\pi_H(\mu) = \pi_H(\mu^{-1}(H))$.

The tuple $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ with $r = (r_i)_{i \in \mathcal{N}}$ and $\pi = (\pi_H)_{H \in \mathcal{D}}$ as above defines a many-to-one matching problem with peer effects and monetary evaluations or simply *matching problem*.

Strict preference profiles are a common assumption in many matching models (cf. Gale and Shapley 1962). Recently, some authors have investigated the link between indifference and inefficiencies, e.g. Erdil and Ergin (2008). In our setting, a *matching problem* is called *strict* if

(i) for every $i \in \mathcal{N}$, all $H_1, H_2 \in \mathcal{D}$, and all $S_1, S_2 \subseteq \mathcal{N}$ with $|S_1| = q_{H_1} |S_2| = q_{H_2}$ and $i \in S_1 \cap S_2$, $r_i(S_1 \setminus \{i\}, H_1) \neq r_i(S_2 \setminus \{i\}, H_2)$ if $S_1 \neq S_2$ or $H_1 \neq H_2$; and

---

[3] More recently Chechlárová (2002) and Chechlárová and Manlove (2005) explore the importance of preference completeness (i.e. whether all agents are acceptable partners to all other agents) and consistency (i.e. when each agent who finds another agent acceptable, is also found acceptable by the former) for the existence of exchange-stable allocations in the context of one-to-one matching problems.

(ii) for every $H \in \mathcal{D}$, all $S_1, S_2 \subseteq \mathcal{N}$ with $|S_1| = |S_2| = q_H, \pi_H(S_1) \neq \pi_H(S_2)$ if $S_1 \neq S_2$.

3 Pareto optimal matchings

In a Pareto optimal matching it is not possible to make any agent better-off by changing their department, in the case of a worker, or their workers, in the case of a department, without making another agent worse-off. The formal definition of a Pareto optimal matching is given below.

Definition 1 Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem. A matching $\mu \in \mathcal{M}$ is Pareto optimal if there is no other matching μ' such that

- $r_i(\mu') \geq r_i(\mu)$ for all $i \in \mathcal{N}$,
- $\pi_H(\mu') \geq \pi_H(\mu)$ for all $H \in \mathcal{D}$, and
- there exists a worker $i \in \mathcal{N}$ such that $r_i(\mu') > r_i(\mu)$ or a department $H \in \mathcal{D}$ such that $\pi_H(\mu') > \pi_H(\mu)$.

We will characterize Pareto optimality of matchings by means of a stability notion called *contractual exchange stability*. Recall that in our set up, departments have fixed capacities and the number of available workers is consistent with the available capacities. Thus, it seems natural to assume that an agents' only possibility for deviation is to perform some switch with other agents on the same side (i.e. workers can exchange departments with other workers and departments can exchange workers with other departments) provided that the exchange meets certain contractual criteria. In particular, if worker i prefers the department and coworkers of worker j in matching μ, and vice versa, then i and j can exchange their places provided that their respective departments and peers are not made worse-off by the swap. Similarly, if two departments, G and H, can generate higher rewards by exchanging a (sub-)set of their workers under μ, then, they can perform the swap provided that none of the involved workers is made worse-off. The description above is limited to two-way exchanges, but we assume that three-way or even more complicated exchanges can be performed as long as the affected parties (peers or departments) do not earn a lower reward as a result of the exchange.

Definition 2 Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem. A matching $\mu \in \mathcal{M}$ is contractually exchange blocked via another matching μ' if either there exists a blocking coalition $\mathcal{S} \subseteq \mathcal{N}$ with $|\mathcal{S}| \geq 2$ such that

(i) for each $i \in \mathcal{S}$ there is a $j \in \mathcal{S}$ with $\mu(i) \neq \mu(j)$ such that $\mu'(i) = \mu(j)$;
(ii) for each $i \in \mathcal{S}, r_i(\mu') > r_i(\mu)$;
(iii) for all $j \in \mathcal{N} \setminus \mathcal{S}, r_j(\mu') \geq r_j(\mu)$;
for all $H \in \mathcal{D}, \pi_H(\mu') \geq \pi_H(\mu)$;

or there exists a blocking coalition $\mathcal{H} \subseteq \mathcal{D}$ with $|\mathcal{H}| \geq 2$ such that

(i') for each $H \in \mathcal{H}$ it holds that $(\mu')^{-1}(H) \neq \mu^{-1}(H)$ and $(\mu')^{-1}(H) \subseteq \cup_{F \in \mathcal{H}} \mu^{-1}(F)$;

(ii') for each $H \in \mathcal{H}$, $\pi_H(\mu') > \pi_H(\mu)$;
(iii') for all $i \in \mathcal{N}$, $r_i(\mu') \geq r_i(\mu)$; for all $F \in \mathcal{D} \setminus \mathcal{H}$, $\pi_F(\mu') \geq \pi_F(\mu)$.

In the above definition, it is required that the blocking coalition contains at least two agents of the same type who agree to a switch and these agents can earn a strictly higher reward as a result of the switch [requirements (ii) and (ii')]. Requirements (iii) and (iii') of the definition, on the other hand, are important only for those agents who are affected by the switches proposed by the blocking coalition. For all other agents, this requirement is automatically satisfied.

Below we introduce the notion of a contractually exchange stable matching.

Definition 3 A matching μ is *contractually exchange stable*, if it cannot be contractually exchange blocked.

Theorem 1 *Let* $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ *be a strict matching problem. A matching* $\mu \in \mathcal{M}$ *is Pareto optimal if and only if it is contractually exchange stable.*

Proof First, let μ be a contractually exchange stable matching. We show that μ is Pareto optimal by contradiction. Suppose that μ is not Pareto optimal. Therefore, there exists a matching μ' such that

1. $r_i(\mu') \geq r_i(\mu)$ for all $i \in \mathcal{N}$,
2. $\pi_H(\mu') \geq \pi_H(\mu)$ for all $H \in \mathcal{D}$, and
3. there exists a worker $i \in \mathcal{N}$ such that $r_i(\mu') > r_i(\mu)$ or a department $H \in \mathcal{D}$ such that $\pi_H(\mu') > \pi_H(\mu)$.

Let $\mathcal{H} = \{H \in \mathcal{D} | (\mu')^{-1}(H) \neq \mu^{-1}(H)\}$. Note that $|\mathcal{H}| \geq 2$ since $\mu' \neq \mu$. By definition of $\mathcal{H}$, we have that

(i') for each $H \in \mathcal{H}$, $(\mu')^{-1}(H) \neq \mu^{-1}(H)$ and $(\mu')^{-1}(H) \subseteq \cup_{F \in \mathcal{H}} \mu^{-1}(F)$.
Since $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ is a strict matching problem and taking (2) above into account, we have that
(ii') for each $H \in \mathcal{H}$, $\pi_H(\mu') > \pi_H(\mu)$.
Moreover, by points (1) and (2) above, we know that
(iii') for all $i \in \mathcal{N}$, $r_i(\mu') \geq r_i(\mu)$; for all $F \in \mathcal{D} \setminus \mathcal{H}$, $\pi_F(\mu') \geq \pi_F(\mu)$.

Therefore, μ can be blocked via μ' by coalition $\mathcal{H}$. This establishes a contradiction to our premise that μ is contractually exchange stable.

Next, let μ be a Pareto optimal matching. We show that μ is contractually exchange stable by contradiction. Suppose that μ is not contractually exchange stable. Then, there exists a matching μ' that contractually exchange blocks μ. We consider two cases: (a) $\mathcal{S}$ is a blocking coalition of workers and (b) $\mathcal{H}$ is a blocking coalition of departments.

(a) Let $\mathcal{S} \subseteq \mathcal{N}$ with $|\mathcal{S}| \geq 2$ be a blocking coalition of workers. From (ii) and (iii) in Definition 2, one readily derives that
 – for all $i \in \mathcal{N}$, $r_i(\mu') \geq r_i(\mu)$ and
 – for all $H \in \mathcal{D}$, $\pi_H(\mu') \geq \pi_H(\mu)$.
Moreover, from (ii) in Definition 2 and $\mathcal{S} \neq \emptyset$, we find that
 – $r_i(\mu') > r_i(\mu)$ for some $i \in \mathcal{N}$.

This establishes a contradiction to our premise that μ is Pareto optimal.

(a) Let $\mathcal{H} \subseteq \mathcal{D}$ with $|\mathcal{H}| \geq 2$ be a blocking coalition of departments. From (ii′) and (iii′) in Definition 2, one readily derives that

 – for all $i \in \mathcal{N}, r_i(\mu') \geq r_i(\mu)$ and
 – for all $H \in \mathcal{D}, \pi_H(\mu') \geq \pi_H(\mu)$.

Moreover, from (ii′) in Definition 2 and $\mathcal{H} \neq \emptyset$, we find that

 – $\pi_H(\mu') > \pi_H(\mu)$ for some $H \in \mathcal{D}$.

This establishes a contradiction to our premise that μ is Pareto optimal. □

The following example illustrates the need for restricting to the class of strict matching problems in Theorem 1.

Example 1 Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem with $\mathcal{N} = \{1, 2, 3\}$ and $\mathcal{D} = \{A, B\}$, capacities $q_A = 1$ and $q_B = 2$; and reward functions

$$
\begin{aligned}
r_1(\emptyset, A) &= 0, & r_1(\{2\}, B) &= 2, & r_1(\{3\}, B) &= 0, \\
r_2(\emptyset, A) &= 1, & r_2(\{1\}, B) &= 2, & r_2(\{3\}, B) &= 0, \\
r_3(\emptyset, A) &= 0, & r_3(\{1\}, B) &= 2, & r_3(\{2\}, B) &= 1;
\end{aligned}
$$

and

$$
\begin{aligned}
\pi_A(\{1\}) &= 1, & \pi_A(\{2\}) &= 1, & \pi_A(\{3\}) &= 3, \\
\pi_B(\{1, 2\}) &= 3, & \pi_B(\{1, 3\}) &= 2, & \pi_B(\{2, 3\}) &= 1.
\end{aligned}
$$

Clearly, $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ is not a strict matching problem since worker 1 has the same reward when he is assigned to department A and when he is assigned to department B with co-worker 3. Also, department A has the same reward when assigned to workers 1 and 2. We show that in this situation it is possible for a matching to be contractually exchange stable, but not Pareto optimal.

Consider the matching μ_1 such that $\mu_1(1) = A, \mu_1(2) = \mu_1(3) = B$. This matching is contractually exchange stable. Any coalition of workers that attempts to block the matching must be either $\{1, 2\}$, or $\{1, 3\}$. First, consider workers 1 and 2: 1 does not want to exchange assignments with 2 because his rewards are the same under μ_1 and under a new matching in which he is matched to B together with 3. Furthermore, matching μ_1 cannot be blocked by workers 1 and 3 because for worker 3 we have that $r_3(\emptyset, A) = 0 < 1 = r_3(\mu_1)$.

Similarly, we can show that μ_1 cannot be contractually exchange blocked by $\mathcal{H} = \{A, B\}$. Departments A and B cannot block μ_1 by exchanging workers 1 and 3 because worker 3 earns lower reward when matched to department A than when matched with 2 to department B. Departments A and B cannot contractually exchange block μ_1 by exchanging workers 1 and 2 because department A earns the same rewards when assigned to workers 1 or 2.

However, the matching μ_1 is not Pareto optimal. To see that, consider matching μ_2 such that $\mu_2(1) = \mu_2(3) = B$ and $\mu_2(2) = A$ with $r_1(\mu_2) = 0 = r_1(\mu_1), r_2(\mu_2) = 1 > 0 = r_2(\mu_1), r_3(\mu_2) = 2 > 1 = r_3(\mu_1), \pi_A(\mu_2) = 1 = \pi_A(\mu_1)$ and $\pi_B(\mu_2) = 2 > 1 = \pi_B(\mu_1)$.

Note that the requirement of a strict matching problem in Theorem 1 is only used to show that every contractually exchange stable matching is Pareto optimal. Therefore, we have the following result.

Proposition 1 *Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem. If a matching $\mu \in \mathcal{M}$ is Pareto optimal, then it is contractually exchange stable.*

4 Total reward maximizing matchings

In the preceding section, the set of Pareto optimal matchings was characterized by means of a stability notion—contractual exchange stability—that presumes that agents can exercise veto power on deviations involving members of their coalition. As discussed in the introduction, the set of Pareto optimal matchings are desirable from a central manager's point of view due to their efficiency. Within the set of Pareto optimal matchings some outcomes may be more desirable than others. In particular, all Pareto optimal matchings can be ranked on the basis of corresponding total joint rewards of all workers and departments. A central manager may thus be interested not only in achieving a Pareto optimal matching, but also in achieving a Pareto optimal matching that corresponds to maximum total reward. Note that we will refer to such matchings as matchings of maximum total reward and give a formal definition of this notion below. The question that this section addresses is what kind of stability notion characterize this subset of the set of Pareto optimal matchings, namely, the set of matchings of maximum total reward.

Definition 4 Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem. A *matching* $\mu \in \mathcal{M}$ is *of maximum total reward* if

$$\sum_{i \in \mathcal{N}} r_i(\mu) + \sum_{H \in \mathcal{D}} \pi_H(\mu) \geq \sum_{i \in \mathcal{N}} r_i(\mu') + \sum_{H \in \mathcal{D}} \pi_H(\mu') \quad \text{for all } \mu' \in \mathcal{M}.$$

Note that every matching μ of maximum total reward is automatically Pareto optimal.

Example 2 Reconsider the matching problem of Example 1. In this problem, there are two Pareto optimal matchings: μ_1 such that $\mu_1(2) = A$ and $\mu_1(1) = \mu_1(3) = B$; and μ_2 such that $\mu_2(3) = A$ and $\mu_2(1) = \mu_2(2) = B$. There is, however, only one matching with a maximum total reward and that is μ_2 as the total reward of μ_1 is 6 and that of μ_2 is 10.

Next, we present a new stability notion, compensational stability, to (later) characterize the set of matchings of maximum total reward. In contrast to the notion of contractual stability, compensation stability exploits the cardinal property of the agents' reward function. Underlying the notion of compensation stability is the additional assumption that deviating agents can use the increase in their reward due to their deviation to compensate agents adversely affected by the deviation.

Definition 5 Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem and let $\mu \in \mathcal{M}$ be a matching. The matching μ is *compensationally stable* if there is no other matching μ' such that

$$\sum_{i \in \mathcal{S} \cup \mathcal{T}} r_i(\mu') + \sum_{H \in \mathcal{F}} \pi_H(\mu') > \sum_{i \in \mathcal{S} \cup \mathcal{T}} r_i(\mu) + \sum_{H \in \mathcal{F}} \pi_H(\mu),$$

where $\mathcal{S} = \{i \in \mathcal{N} \mid \mu'(i) \neq \mu(i)\}$, $\mathcal{T} = \{i \in \mathcal{N} \setminus \mathcal{S} \mid \mu^{-1}(\mu(i)) \cap \mathcal{S} \neq \emptyset\}$, and $\mathcal{F} = \{H \in \mathcal{D} \mid (\mu')^{-1}(H) \neq \mu^{-1}(H)\}$, i.e. $\mathcal{S}$ is the set of workers that are assigned to a different department, $\mathcal{T}$ is the set of workers that do not change department but change peers, and $\mathcal{F}$ is the group of departments that get a different set of workers.

Theorem 2 *Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem and let $\mu \in \mathcal{M}$ be a matching. Then, μ is of maximum total reward if and only if it is compensationally stable.*

Proof First, let μ be a compensationally stable matching. We show that μ is of maximum total reward by contradiction. Suppose that μ is not of maximum total reward. Then, there is a matching μ' such that

$$\sum_{i \in \mathcal{N}} r_i(\mu') + \sum_{H \in \mathcal{D}} \pi_H(\mu') > \sum_{i \in \mathcal{N}} r_i(\mu) + \sum_{H \in \mathcal{D}} \pi_H(\mu). \tag{1}$$

With $\mathcal{S} = \{i \in \mathcal{N} \mid \mu'(i) \neq \mu(i)\}$, $\mathcal{T} = \{i \in \mathcal{N} \setminus \mathcal{S} \mid \mu^{-1}(\mu(i)) \cap \mathcal{S} \neq \emptyset\}$, and $\mathcal{F} = \{H \in \mathcal{D} \mid (\mu')^{-1}(H) \neq \mu^{-1}(H)\}$, it follows that

$$r_i(\mu') = r_i(\mu) \quad \text{for all } i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T}) \text{ and} \tag{2}$$
$$\pi_H(\mu') = \pi_H(\mu) \quad \text{for all } H \in \mathcal{D} \setminus \mathcal{F}. \tag{3}$$

Then,

$$
\begin{aligned}
\sum_{i \in \mathcal{S} \cup \mathcal{T}} r_i(\mu') + \sum_{H \in \mathcal{F}} \pi_H(\mu') &= \sum_{i \in \mathcal{N}} r_i(\mu') + \sum_{H \in \mathcal{D}} \pi_H(\mu') \\
&\quad - \sum_{i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T})} r_i(\mu') - \sum_{H \in \mathcal{D} \setminus \mathcal{F}} \pi_H(\mu') \\
&= \sum_{i \in \mathcal{N}} r_i(\mu') + \sum_{H \in \mathcal{D}} \pi_H(\mu') \\
&\quad - \sum_{i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T})} r_i(\mu) - \sum_{H \in \mathcal{D} \setminus \mathcal{F}} \pi_H(\mu) \\
&> \sum_{i \in \mathcal{N}} r_i(\mu) + \sum_{H \in \mathcal{D}} \pi_H(\mu) \\
&\quad - \sum_{i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T})} r_i(\mu) - \sum_{H \in \mathcal{D} \setminus \mathcal{F}} \pi_H(\mu) \\
&= \sum_{i \in \mathcal{S} \cup \mathcal{T}} r_i(\mu) + \sum_{H \in \mathcal{F}} \pi_H(\mu)
\end{aligned}
$$

where the second equality follows by Eqs. (2) and (3), and the inequality follows by Eq. (1). This establishes a contradiction to our premise that μ is compensationally stable.

Next, let μ be a matching of maximum total reward. We show that μ is compensationally stable by contradiction. Suppose that μ is not compensationally stable. Then, there is a matching μ' such that

$$\sum_{i \in \mathcal{T} \cup \mathcal{S}} r_i(\mu') + \sum_{H \in \mathcal{F}} \pi_H(\mu') > \sum_{i \in \mathcal{T} \cup \mathcal{S}} r_i(\mu) + \sum_{H \in \mathcal{F}} \pi_H(\mu), \tag{4}$$

where $\mathcal{S} = \{i \in \mathcal{N} \mid \mu'(i) \neq \mu(i)\}$, $\mathcal{T} = \{i \in \mathcal{N} \setminus \mathcal{S} \mid \mu^{-1}(\mu(i)) \cap \mathcal{S} \neq \emptyset\}$, and $\mathcal{F} = \{H \in \mathcal{D} \mid (\mu')^{-1}(H) \neq \mu^{-1}(H)\}$. It follows that

$$\pi_H(\mu') = \pi_H(\mu) \text{ for all } H \in \mathcal{D} \setminus \mathcal{F} \text{ and} \tag{5}$$
$$r_i(\mu') = r_i(\mu) \text{ for all } i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T}). \tag{6}$$

Then,

$$\begin{aligned}
\sum_{i \in \mathcal{N}} r_i(\mu') + \sum_{H \in \mathcal{D}} \pi_H(\mu') &= \sum_{i \in \mathcal{T} \cup \mathcal{S}} r_i(\mu') + \sum_{H \in \mathcal{F}} \pi_H(\mu') \\
&\quad + \sum_{i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T})} r_i(\mu') + \sum_{H \in \mathcal{D} \setminus \mathcal{F}} \pi_H(\mu') \\
&= \sum_{i \in \mathcal{T} \cup \mathcal{S}} r_i(\mu') + \sum_{H \in \mathcal{F}} \pi_H(\mu') \\
&\quad + \sum_{i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T})} r_i(\mu) + \sum_{H \in \mathcal{D} \setminus \mathcal{F}} \pi_H(\mu) \\
&> \sum_{i \in \mathcal{T} \cup \mathcal{S}} r_i(\mu) + \sum_{H \in \mathcal{F}} \pi_F(\mu) \\
&\quad + \sum_{i \in \mathcal{N} \setminus (\mathcal{S} \cup \mathcal{T})} r_i(\mu) + \sum_{H \in \mathcal{D} \setminus \mathcal{F}} \pi_H(\mu) \\
&= \sum_{i \in \mathcal{N}} r_i(\mu) + \sum_{H \in \mathcal{D}} \pi_H(\mu)
\end{aligned}$$

where the second equality follows by Eqs. (5) and (6), and the inequality follows by Eq. (4). This establishes a contradiction to our premise that μ is of maximum total reward. □

5 From an initial matching to a maximum total reward matching

In this section, we analyze situations in which there is an initial matching of workers to departments which does not generate maximum total reward. Thus, there exists a possible restructuring where agents are reassigned by means of a matching of maximum total reward. The question that arises is how to compensate those workers or

departments that are worse off in the new situation. This question will be analyzed using a cooperative matching game with transferable utility.

A *cooperative (transferable utility) game* in characteristic function form is a pair (N, v) where N is a finite set of players and $v : 2^N \to \mathbb{R}$ satisfying $v(\emptyset) = 0$. In general, $v(\mathcal{C})$ represents the maximal joint reward that coalition $\mathcal{C} \in 2^N$ can obtain when its members cooperate in an optimal way, without help of the players in $N \setminus \mathcal{C}$.

The *core* of a game (N, v) is defined by

$$\text{Core}(v) = \left\{ x \in \mathbb{R}^N | \sum_{i \in N} x_i = v(N), \sum_{i \in \mathcal{C}} x_i \geq v(\mathcal{C}) \text{ for all } \mathcal{C} \in 2^N \right\},$$

i.e. the core is the set of efficient allocations of $v(N)$ to which no coalition can reasonably object. An important subclass of games with nonempty core is the class of convex games (see Shapley (1971)). A game (N, v) is said to be *convex* if $v(\mathcal{C}^1 \cup \{i\}) - v(\mathcal{C}^1) \leq v(\mathcal{C}^2 \cup \{i\}) - v(\mathcal{C}^2)$ for every $i \in N$ and every $\mathcal{C}^1 \subseteq \mathcal{C}^2 \subseteq N \setminus \{i\}$.

Before starting with the formal analysis of compensation in matching problems with an initial matching, we provide an illustrative example.

Example 3 Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem with $\mathcal{N} = \{1, 2, 3, 4\}$ and $\mathcal{D} = \{A, B\}$, capacities $q_A = 2$ and $q_B = 2$, and reward functions

$$\begin{aligned}
&r_1(\{2\}, A) = r_1(\{2\}, B) = 20, \quad r_1(\{3\}, A) = r_1(\{3\}, B) = 7, \\
&r_1(\{4\}, A) = r_1(\{4\}, B) = 14, \\
&r_2(\{1\}, A) = r_2(\{1\}, B) = 10, \quad r_2(\{3\}, A) = r_2(\{3\}, B) = 3, \\
&r_2(\{4\}, A) = r_2(\{4\}, B) = 9, \\
&r_3(\{1\}, A) = r_3(\{1\}, B) = 15, \quad r_3(\{2\}, A) = r_3(\{2\}, B) = 7, \\
&r_3(\{4\}, A) = r_3(\{4\}, B) = 10, \\
&r_4(\{1\}, A) = r_4(\{1\}, B) = 14, \quad r_4(\{2\}, A) = r_4(\{2\}, B) = 19, \\
&r_4(\{3\}, A) = r_4(\{3\}, B) = 10
\end{aligned}$$

and

$$\pi_A(\mathcal{S}) = \pi_B(\mathcal{S}) = 10$$

for every $\mathcal{S} \in 2^{\mathcal{N}}$ with $|\mathcal{S}| = 2$.

Assume that the initial matching μ_0 is given by $\mu_0(1) = \mu_0(4) = A$ and $\mu_0(2) = \mu_0(3) = B$. In this case, the initial situation has a total reward of 58 with $r_1(\mu_0) = 14, r_2(\mu_0) = 3, r_3(\mu_0) = 7, r_4(\mu_0) = 14, \pi_A(\mu_0) = 10$ and $\pi_B(\mu_0) = 10$, while the maximum social reward is 70 which can be achieved, for example, by $\hat{\mu}$ with $\hat{\mu}(1) = \hat{\mu}(2) = A$ and $\hat{\mu}(3) = \hat{\mu}(4) = B$ with $r_1(\hat{\mu}) = 20, r_2(\hat{\mu}) = 10, r_3(\hat{\mu}) = 10, r_4(\hat{\mu}) = 10, \pi_A(\hat{\mu}) = 10$ and $\pi_B(\hat{\mu}) = 10$. Note that $r_4(\hat{\mu}) = 10 < r_4(\mu_0) = 14$.

Given a matching problem $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ and an arbitrary initial matching μ_0, we are interested in studying adequate compensations among the workers and departments

when they come together to improve their joint situation by adopting a matching of maximum total reward. To do this, we define a corresponding cooperative game whose core elements provide stable allocations of the extra rewards obtained by means of cooperation.

Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem with constraints and let μ_0 be an initial matching. We denote by $\mathcal{M}^{\text{max}}$ the set of matchings of maximal total reward, i.e.

$$\mathcal{M}^{\text{max}} = \{\hat{\mu} \in \mathcal{M} \mid \hat{\mu} \text{ is of maximal total reward}\}.$$

For $i \in \mathcal{N}$ and $\hat{\mu} \in \mathcal{M}^{\text{max}}$, we denote by $d_i(\hat{\mu})$ the difference of reward for agent $i \in \mathcal{N}$ when going from μ_0 to $\hat{\mu}$, i.e. $d_i(\hat{\mu}) = r_i(\hat{\mu}) - r_i(\mu_0)$; we set $d_i^+(\hat{\mu}) = \max\{0, r_i(\hat{\mu}) - r_i(\mu_0)\}$ and $d_i^-(\hat{\mu}) = -\min\{0, r_i(\hat{\mu}) - r_i(\mu_0)\}$. Note that for every $\hat{\mu} \in \mathcal{M}^{\text{max}}$ it follows $d_i(\hat{\mu}) = d_i^+(\hat{\mu}) - d_i^-(\hat{\mu})$. We analogously define $d_H(\hat{\mu}) = \pi_H(\hat{\mu}) - \pi_H(\mu_0)$, and set $d_H^+(\hat{\mu}) = \max\{0, \pi_H(\hat{\mu}) - \pi_H(\mu_0)\}$, and $d_H^-(\hat{\mu}) = -\min\{0, \pi_H(\hat{\mu}) - \pi_H(\mu_0)\}$ for every $H \in \mathcal{D}$ and every $\hat{\mu} \in \mathcal{M}^{\text{max}}$.

Definition 6 Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem and let μ_0 be an initial matching. The corresponding *compensation matching game*, $(\mathcal{N} \cup \mathcal{D}, v)$, is defined by

$$v(\mathcal{C}) = \min_{\hat{\mu} \in \mathcal{M}^{\text{max}}} \left\{ v^{\hat{\mu}}(\mathcal{C}) \right\} \tag{7}$$

for every $\mathcal{C} \subset \mathcal{N} \cup \mathcal{D}$, where

$$v^{\hat{\mu}}(\mathcal{C}) = \max \left\{ 0, \sum_{i \in \mathcal{C} \cap \mathcal{N}} d_i^+(\hat{\mu}) + \sum_{H \in \mathcal{C} \cap \mathcal{D}} d_H^+(\hat{\mu}) - \sum_{i \in \mathcal{N}} d_i^-(\hat{\mu}) - \sum_{H \in \mathcal{D}} d_H^-(\hat{\mu}) \right\}. \tag{8}$$

Note that given a matching with maximum total reward $\hat{\mu}$, we conservatively assume that a coalition $\mathcal{C}$ is required to compensate *all* agents that suffer from changing the initial situation. Being conservative, it is also assumed that a most disadvantageous (from the perspective of $\mathcal{C}$) matching of maximum total reward will be adopted. Note that $v(\mathcal{N} \cup \mathcal{D})$ equals the increase in total rewards achievable by adopting any matching with maximum total reward.

The following example illustrates the computation of matching games with constraints.

Example 4 Reconsider the matching problem $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ of Example 3 and let μ_0 be given by $\mu_0(1) = \mu_0(4) = A$ and $\mu_0(2) = \mu_0(3) = B$, with a social reward of 58. There are four matchings with maximum total reward, namely

$$\begin{aligned}
&\hat{\mu}_1: \hat{\mu}_1(1) = \hat{\mu}_1(2) = A, &&\hat{\mu}_1(3) = \hat{\mu}_1(4) = B, \\
&\hat{\mu}_2: \hat{\mu}_2(3) = \hat{\mu}_2(4) = A, &&\hat{\mu}_2(1) = \hat{\mu}_2(2) = B, \\
&\hat{\mu}_3: \hat{\mu}_3(1) = \hat{\mu}_3(3) = A, &&\hat{\mu}_3(2) = \hat{\mu}_3(4) = B, \\
&\hat{\mu}_4: \hat{\mu}_4(2) = \hat{\mu}_4(4) = A, &&\hat{\mu}_4(1) = \hat{\mu}_4(3) = B.
\end{aligned}$$

Table 1 Individual extra rewards and losses for the agents in Example 4

$\hat{\mu}$	$d_1(\hat{\mu})$	$d_2(\hat{\mu})$	$d_3(\hat{\mu})$	$d_4(\hat{\mu})$
$\hat{\mu}_1$	6	7	3	−4
$\hat{\mu}_2$	6	7	3	−4
$\hat{\mu}_3$	−7	6	8	5
$\hat{\mu}_4$	−7	6	8	5
$\hat{\mu}$	$d_1^+(\hat{\mu})$	$d_2^+(\hat{\mu})$	$d_3^+(\hat{\mu})$	$d_4^+(\hat{\mu})$
$\hat{\mu}_1$	6	7	3	0
$\hat{\mu}_2$	6	7	3	0
$\hat{\mu}_3$	0	6	8	5
$\hat{\mu}_4$	0	6	8	5
$\hat{\mu}$	$d_1^-(\hat{\mu})$	$d_2^-(\hat{\mu})$	$d_3^-(\hat{\mu})$	$d_4^-(\hat{\mu})$
$\hat{\mu}_1$	0	0	0	4
$\hat{\mu}_2$	0	0	0	4
$\hat{\mu}_3$	7	0	0	0
$\hat{\mu}_4$	7	0	0	0

The maximum total reward for this problem is 70, therefore there is an extra reward of 12 when going from the initial matching to one of maximum total reward.

The corresponding extra rewards for the departments when going from the initial matching to a matching with maximum total reward is always 0 since they have constant reward functions. The individual extra rewards and losses for the workers when going from the initial matching to a matching of maximum total reward are provided in Table 1.

Since the departments are indifferent between all matching functions, we can omit them from the study of compensations and the player set $\mathcal{N} \cup \mathcal{D}$ can be restricted to $\mathcal{N}$ only. The value of coalition $\{1, 2, 3\}$ in the compensation matching game is explained below. All coalitional values are given in Table 2.

$$
\begin{aligned}
v(\{1, 2, 3\}) &= \min_{l \in \{1,2,3,4\}} \left\{ v^{\hat{\mu}_l}(\{1, 2, 3\}) \right\} \\
&= \min_{l \in \{1,2,3,4\}} \left\{ \max \left\{ 0, \sum_{i=1}^{3} d_i^+(\hat{\mu}_l) - \sum_{i=1}^{4} d_i^-(\hat{\mu}_l) \right\} \right\} \\
&= \min \{ \max \{0, 16 - 4\}, \max \{0, 16 - 4\}, \max \{0, 14 - 7\}, \\
&\quad \max \{0, 14 - 7\} \} \\
&= \min \{12, 12, 7, 7\} \\
&= 7.
\end{aligned}
$$

Note that the core of this game is not empty since, for instance, $(0, 0, 7, 5) \in$ Core(v). Here, the core element $(0, 0, 7, 5)$ can be interpreted as follows: when choos-

Table 2 Coalitional values in Example 4

$\mathcal{C}$	{1}	{2}	{3}	{4}	{1,2}	{1,3}	{1,4}	{2,3}	{2,4}	{3,4}	{1,2,3}	{1,2,4}	{1,3,4}	{2,3,4}	$\mathcal{N}$
$v(\mathcal{C})$	0	0	0	0	0	1	0	6	3	0	7	4	5	6	12

ing matching $\hat{\mu}_3$, worker 2 is helped by worker 3 to fully compensate worker 1 with 6 and 1, respectively.

In addition, note that $(\mathcal{N}, v)$ is not convex since for $i = 4$, $\mathcal{C}^1 = \{2\}$, and $\mathcal{C}^2 = \{2, 3\}$ we have

$$v(\mathcal{C}^1 \cup \{i\}) - v(\mathcal{C}^1) = 3 - 0 > 6 - 6 = v(\mathcal{C}^2 \cup \{i\}) - v(\mathcal{C}^2).$$

It turns out that each compensation matching game has a nonempty core. This can be shown using a relation between compensation matching games and bankruptcy games. A *bankruptcy problem* is defined by a tuple (N, E, c) where $N = \{1, \ldots, n\}$ is the set of agents, E is the estate that must be shared among the agents, and $c \in \mathbb{R}^N$ is the vector of claims of the agents satisfying $\sum_{i \in N} c_i \geq E$. Bankruptcy problems have being studied from a game theoretical viewpoint in O'Neill (1982). Given a bankruptcy problem (N, E, c), the corresponding *bankruptcy game*, $(N, v_{(N,E,c)})$, is defined by

$$v_{(N,E,c)}(\mathcal{C}) = \max \left\{ 0, E - \sum_{i \in N \setminus \mathcal{C}} c_i \right\}$$

for every $\mathcal{C} \subset N$. In Curiel et al. (1987) it is shown that bankruptcy games are convex.

Lemma 1 *Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem, let μ_0 be an initial matching function, and let $\hat{\mu} \in \mathcal{M}^{max}$. Then, $(\mathcal{N} \cup \mathcal{D}, v^{\hat{\mu}})$ is a bankruptcy game.*

Proof Define $E(\hat{\mu}) = \sum_{i \in \mathcal{N}} d_i(\hat{\mu}) + \sum_{H \in \mathcal{D}} d_H(\hat{\mu})$, $c_i = d_i^+(\hat{\mu})$ for every $i \in \mathcal{N}$, and $c_H = d_H^+(\hat{\mu})$ for every $H \in \mathcal{D}$. We will prove that $(\mathcal{N} \cup \mathcal{D}, E(\hat{\mu}), c)$ is a bankruptcy problem with the associated bankruptcy game $v_{(\mathcal{N} \cup \mathcal{D}, E(\hat{\mu}), c)}$ equal to $v^{\hat{\mu}}$.

Note that $(\mathcal{N} \cup \mathcal{D}, E(\hat{\mu}), c)$ is a bankruptcy problem since

$$E(\hat{\mu}) = \sum_{i \in \mathcal{N}} d_i(\hat{\mu}) + \sum_{H \in \mathcal{D}} d_H(\hat{\mu}) = \sum_{i \in \mathcal{N}} d_i^+(\hat{\mu})$$
$$- \sum_{i \in \mathcal{N}} d_i^-(\hat{\mu}) + \sum_{H \in \mathcal{D}} d_H^+(\hat{\mu}) - \sum_{H \in \mathcal{D}} d_H^-(\hat{\mu})$$
$$\leq \sum_{i \in \mathcal{N}} d_i^+(\hat{\mu}) + \sum_{H \in \mathcal{D}} d_H^+(\hat{\mu}) = \sum_{i \in \mathcal{N}} c_i + \sum_{H \in \mathcal{D}} c_H$$

by definition of $d_i^-(\hat{\mu})$ and $d_H^-(\hat{\mu})$. Moreover, for $\mathcal{C} \subset \mathcal{N} \cup \mathcal{D}$, we have

$$v^{\hat{\mu}}(\mathcal{C}) = \max \left\{ 0, \sum_{i \in \mathcal{C} \cap \mathcal{N}} d_i^+(\hat{\mu}) + \sum_{H \in \mathcal{C} \cap \mathcal{D}} d_H^+(\hat{\mu}) - \sum_{i \in \mathcal{N}} d_i^-(\hat{\mu}) - \sum_{H \in \mathcal{D}} d_H^-(\hat{\mu}) \right\}$$

$$= \max \left\{ 0, \sum_{i \in \mathcal{N}} d_i^+(\hat{\mu}) + \sum_{H \in \mathcal{D}} d_H^+(\hat{\mu}) - \sum_{i \in \mathcal{N}} d_i^-(\hat{\mu}) - \sum_{H \in \mathcal{D}} d_H^-(\hat{\mu}) \right.$$

$$\left. - \sum_{i \in \mathcal{N} \backslash (\mathcal{C} \cap \mathcal{N})} d_i^+(\hat{\mu}) - \sum_{H \in \mathcal{D} \backslash (\mathcal{C} \cap \mathcal{D})} d_H^+(\hat{\mu}) \right\}$$

$$= \max \left\{ 0, \sum_{i \in \mathcal{N}} d_i(\hat{\mu}) + \sum_{H \in \mathcal{D}} d_H(\hat{\mu}) - \sum_{i \in \mathcal{N} \backslash (\mathcal{C} \cap \mathcal{N})} d_i^+(\hat{\mu}) - \sum_{H \in \mathcal{D} \backslash (\mathcal{C} \cap \mathcal{D})} d_H^+(\hat{\mu}) \right\}$$

$$= \max \left\{ 0, E(\hat{\mu}) - \sum_{i \in \mathcal{N} \backslash (\mathcal{C} \cap \mathcal{N})} c_i - \sum_{H \in \mathcal{D} \backslash (\mathcal{C} \cap \mathcal{D})} c_H \right\}$$

$$= v_{(\mathcal{N} \cup \mathcal{D}, E(\hat{\mu}), c)}(\mathcal{C}).$$

$\square$

Theorem 3 *Compensation matching games have a nonempty core.*

Proof Let $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ be a matching problem, let μ_0 be an initial matching function, and let $(\mathcal{N} \cup \mathcal{D}, v)$ be the corresponding compensation matching game. By Lemma 1, we know that v is the minimum of a finite number of bankruptcy games with the value of the grand coalition equal for all games. Clearly, this implies that $\text{Core}(v^{\hat{\mu}}) \subseteq \text{Core}(v)$ for every $\hat{\mu} \in \mathcal{M}^{\max}$. Because bankruptcy games are convex, they have a nonempty core and, therefore, $\text{Core}(v^{\hat{\mu}}) \neq \emptyset$ for every $\hat{\mu} \in \mathcal{M}^{\max}$, which implies $\text{Core}(v) \neq \emptyset$. $\square$

The following example illustrates the construction of a bankruptcy game given a matching problem and an initial matching and the convexity property of these games.

Example 5 Reconsider the matching problem $(\mathcal{N}, \mathcal{D}, q, r, \pi)$ of Example 3 and the initial matching μ_0 given by $\mu_0(1) = \mu_0(4) = A$ and $\mu_0(2) = \mu_0(3) = B$ discussed in Example 4. As shown in Example 4, there are four matchings with maximum total reward. Here, we will consider matching $\hat{\mu}_1$ given by $\hat{\mu}_1(1) = \hat{\mu}_1(2) = A$ and $\hat{\mu}_1(3) = \hat{\mu}_1(4) = B$. The first row of Table 1 shows the extra rewards and losses for the agents when going from the initial matching, μ_0, to the matching of maximum total reward $\hat{\mu}_1$. Recall that the corresponding extra rewards for the institutions is always 0 since they have constant reward functions in the example.

To construct the bankruptcy game $(\mathcal{N} \cup \mathcal{D}, v^{\hat{\mu}})$, we use Equation 8 and compute the value $v^{\hat{\mu}_1}(\mathcal{C})$ for every $\mathcal{C} \subset \mathcal{N} \cup \mathcal{D}$. As an example, the value of coalition $\{1, 2\}$ is computed below.

$$v^{\hat{\mu}_1}(\{1, 2\}) = \max \left\{ 0, \sum_{i \in \{1,2\}} d_i^+(\hat{\mu}_1) - \sum_{i \in \{1,2,3,4\}} d_i^-(\hat{\mu}_1) \right\}$$

$$= \max \left\{ 0, \sum_{i=1}^{2} d_i^+(\hat{\mu}_l) - \sum_{i=1}^{4} d_i^-(\hat{\mu}_l) \right\}$$

Table 3 Coalitional values in Example 5

$\mathcal{C}$	{1}	{2}	{3}	{4}	{1,2}	{1,3}	{1,4}	{2,3}	{2,4}	{3,4}	{1,2,3}	{1,2,4}	{1,3,4}	{2,3,4}	$\mathcal{N}$
v($\mathcal{C}$)	2	3	0	0	9	5	2	6	3	0	12	9	5	6	12

$$= \max \{0, 13 - 4\}$$
$$= 9.$$

All coalitional values are given in Table 3.

Note that the core of this game is not empty since, for instance, $(4, 5, 3, 0) \in \text{Core}(v^{\hat{\mu}_1})$. To see that $(\mathcal{N} \cup \mathcal{D}, v^{\hat{\mu}})$ is convex, consider as an example coalitions $\mathcal{C}_1 = \{1, 2\}$ and $\mathcal{C}_2 = \{1, 2, 3\}$ and $i = 4$. We have $v(\mathcal{C}^1 \cup \{i\}) - v(\mathcal{C}^1) = 9 - 9 = 12 - 12 = v(\mathcal{C}^2 \cup \{i\}) - v(\mathcal{C}^2)$. Similarly, one can check that the convexity property holds with respect to every $i \in \{1, 2, 3, 4\}$ and every $\mathcal{C}^1 \subseteq \mathcal{C}^2 \subseteq \{1, 2, 3, 4\} \setminus \{i\}$.

Theorem 3 tells us that we can find a compensation schedule such that a centralized restructuring resulting in a matching of maximum total reward of the organization will be ex-post stable, in the sense that there will be no subset of workers or departments who can by deviation obtain a higher reward. Still, the core of matching games may contain many possible allocations, thus, the question remains what allocation will be selected. The following example illustrates how such a selection could be conducted.

Example 6 Reconsider the compensation matching game of Example 4. Recall that there are four matchings with maximum total reward:

$$\hat{\mu}_1 \colon \hat{\mu}_1(1) = \hat{\mu}_1(2) = A, \quad \hat{\mu}_1(3) = \hat{\mu}_1(4) = B,$$
$$\hat{\mu}_2 \colon \hat{\mu}_2(3) = \hat{\mu}_2(4) = A, \quad \hat{\mu}_2(1) = \hat{\mu}_2(2) = B,$$
$$\hat{\mu}_3 \colon \hat{\mu}_3(1) = \hat{\mu}_3(3) = A, \quad \hat{\mu}_3(2) = \hat{\mu}_3(4) = B,$$
$$\hat{\mu}_4 \colon \hat{\mu}_4(2) = \hat{\mu}_4(4) = A, \quad \hat{\mu}_4(1) = \hat{\mu}_4(3) = B.$$

and that there is an extra reward of 12 when going from the initial matching to one of maximum total reward. Since $v(\mathcal{C}) = \min \left\{ v^{\hat{\mu}_1}(\mathcal{C}), v^{\hat{\mu}_2}(\mathcal{C}), v^{\hat{\mu}_3}(\mathcal{C}), v^{\hat{\mu}_4}(\mathcal{C}) \right\}$ with $v^{\hat{\mu}_1}(\mathcal{N} \cup \mathcal{D}) = v^{\hat{\mu}_2}(\mathcal{N} \cup \mathcal{D}) = v^{\hat{\mu}_3}(\mathcal{N} \cup \mathcal{D}) = v^{\hat{\mu}_4}(\mathcal{N} \cup \mathcal{D})$, we have that $\text{Core}(v^{\hat{\mu}_l}) \subset \text{Core}(v)$. Moreover, since $(\mathcal{N} \cup \mathcal{D}, v^{\hat{\mu}_l})$ is a bankruptcy game for every $l \in \{1, 2, 3, 4\}$ by Lemma 1, the allocation given by any bankruptcy rule applied to the bankruptcy problem associated to $\hat{\mu}_l$, $l \in \{1, 2, 3, 4\}$, is in the core of the compensation matching game, as well as any convex combination of those allocations. As an illustration, Table 4 gives the allocations obtained by applying the proportional rule (Prop), the adjusted proportional rule (AProp), the constrained equal awards rule (CEA), and the Talmud rule (Tal) to the bankruptcy problems as defined in the proof of Lemma 1. For a description of these rules see Thomson (2003).

If the implemented matching of maximal total reward is known a priori, one can use a specific bankruptcy rule, for example, the proportional rule, in the corresponding bankruptcy problem associated to that selected matching. Otherwise, if a probability

Table 4 Bankruptcy rules for compensation matching games in Example 6

	$\hat{\mu}_1, \hat{\mu}_2 : (\mathcal{N}, 12, (6, 7, 3, 0))$	$\hat{\mu}_3, \hat{\mu}_4 : (\mathcal{N}, 12, (0, 6, 8, 5))$
Prop(N, E, c)	$\left(4\frac{1}{2}, 5\frac{1}{4}, 2\frac{1}{4}, 0\right)$	$\left(0, 3\frac{15}{19}, 5\frac{1}{19}, 3\frac{3}{19}\right)$
AProp(N, E, c)	$\left(4\frac{6}{11}, 5\frac{6}{11}, 1\frac{10}{11}, 0\right)$	$\left(0, 3\frac{2}{3}, 5\frac{5}{18}, 3\frac{1}{18}\right)$
CEA(N, E, c)	$(4.5, 4.5, 3, 0)$	$(0, 4, 4, 4)$
Tal(N, E, c)	$\left(4\frac{2}{3}, 5\frac{2}{3}, 1\frac{2}{3}, 0\right)$	$\left(0, 3\frac{2}{3}, 5\frac{2}{3}, 2\frac{2}{3}\right)$

distribution over the possible acceptance of a matching of maximal total reward is known a priori, a possible allocation can be the expected allocation given by a specific bankruptcy rule.

References

Alcalde, J. (1995). Exchange-proofness or divorce-proofness? *Stability in one-sided matching markets, Economic Design*, *1*, 275–287.

Bogomolnaia, A., & Jackson, M. (2002). The stability of hedonic coalition structures. *Games and Economic Behavior*, *38*, 201–230.

Chechlárová, K. (2002). On the complexity of exchange-stable roommates. *Discrete Applied Mathematics*, *116*, 279–287.

Chechlárová, K., & Manlove, D. (2005). Exchange-stable marriage problem. *Discrete Applied Mathematics*, *152*, 109–122.

Curiel, I. J., Maschler, M., & Tijs, S. (1987). Bankruptcy games. *Zeitschrift für Operations Research*, *31*, A 143–A 159.

Dutta, B., & Massó, J. (1997). Stability of matchings when individuals have preferences over colleagues. *Journal of Economic Theory*, *75*, 464–475.

Erdil, A., & Ergin, H. (2008). What's the matter with tie-breaking? *Improving efficiency in school choice, American Economic Review*, *98*(3), 669–689.

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

Knuth, H. W. (1955). The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, *2*, 83–87.

Koopmans, T. C., & Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrica*, *25*, 53–76.

Lazarova, E., Borm, P., & van Velzen, B. (2011). Coalitional games and contracts based on individual deviations. *Top*, *19*, 507–520.

O'Neill, B. (1982). A problem of rights arbitration from the Talmud. *Mathematical Social Sciences*, *2*, 345–371.

Pentico, D. W. (2007). Assignment problems: A golden anniversary survey. *European Journal of Operational Research*, *176*, 774–793.

Pycia, M. (2012). Stability and preference alignment in matching and coalition formation. *Econometrica*, *80*, 323–362.

Shapley, L. S. (1971). Cores of convex games. *International Journal of Game Theory*, *1*, 11–26.

Shapley, L. S., & Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, *1*, 111130.

Sotomayor, M. (1999). The lattice structure of the set of stable outcomes of the multiple partners assignment game. *International Journal of Game Theory*, *28*, 567–583.

Thomson, W. (2003). Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey. *Mathematical Social Sciences*, *45*, 249–297.