J. Econ. Theory 216 (2024) 105780

Contents lists available at ScienceDirect

Journal of Economic Theory

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

Unilateral stability in matching problems ☆

Michael Richter [a,b,*], Ariel Rubinstein [c,d]

[a] Baruch College, City University of New York, United States of America
[b] Royal Holloway, University of London, United Kingdom of Great Britain and Northern Ireland
[c] School of Economics, Tel Aviv University, Israel
[d] Department of Economics, New York University, United States of America

ARTICLE INFO

JEL classification:
C78

Keywords:
Matching problems
Pairwise stability
Unilateral stability

ABSTRACT

The canonical solution concept used in matching problems is pairwise stability, whose premise is that harmony is disrupted by any two agents intentionally leaving their partners to be with each other. We instead focus on scenarios in which harmony is disrupted merely by a single agent unilaterally initiating contact with a member of a different pair, whether or not his approach is reciprocated. A variety of solution concepts are proposed in which taboos, status, or power systematically limit such initiatives in order to achieve harmony.

1. Introduction

Matching problems form a class of iconic models in economic theory beginning with Gale and Shapley (1962). Classically, the threat to stability in these models is two members agreeing to leave their partners and form a new match. In contrast, we focus on scenarios in which instability arises from a single agent unilaterally approaching another in a different pair.

In the model that we work with, N is an even-numbered population of n agents who must partition themselves into pairs. Each agent i has strict preferences $\succ^i$ over the other agents. This formalization is the *one-sided matching problem*, but it also contains the

☆ The Orchard Tea Garden in Cambridge is the birthplace of this paper. We are grateful to Áron Tóbiás for his very useful comments. Our thanks to the editor of this journal and two referees for very helpful suggestions.

* Corresponding author.
E-mail address: michael.dan.richter@gmail.com (M. Richter).

https://doi.org/10.1016/j.jet.2023.105780
Received 22 August 2023; final version received 16 December 2023; Accepted 17 December 2023

Available online 28 December 2023

two-sided matching problem in which the agents are partitioned into two equally-sized groups, N_1 and N_2, and each agent ranks all agents from his own group below all agents in the other group.

A *pairing* is a profile $(x^i)_{i\in N}$ that specifies for every agent i a partner $x^i \neq i$, such that for any two agents i and j, if $x^i = j$ then $x^j = i$. We write $i \leftrightarrow j$ to denote a match between i and j. A pairing is said to be *pairwise stable* if there are no two agents in different pairs, each of whom prefers the other to his current partner. As established in Gale and Shapley (1962), a pairwise-stable pairing exists for the two-sided matching problem, but not necessarily for the one-sided matching problem. However, we are not motivated by the issue of existence of pairwise stable pairings. Rather, our motivation is conceptual; we seek to model situations where the overall harmony of the system can be jeopardized by a unilateral move of a single agent rather than a joint move of a pair of agents.

Here is the situation we have in mind: All agents are paired up. Any agent can initiate contact with any other and propose that they form a new match. Social harmony is disturbed not just by the actual formation of this proposed match, but rather by the mere act of one agent approaching another with such a proposal. The result is instability since once an approach is made, a scandal ensues: the abandoned partner feels resentment, and whether the approach is reciprocated or not, bad sentiments spread throughout society.

In other words, a society can be destabilized by agent A approaching B and expressing his desire for B to abandon B's partner and match with him instead. There are several possible motives for A's approach: He might know that B also prefers him over B's current partner (which is the premise of pairwise stability). Alternatively, A might not know B's preferences and either simply tries his luck or hopes that even if B does not currently prefer him over his partner, B will feel flattered by A's approach and may change his opinion.

Apart from very rare preference configurations, it is not possible for every agent to be paired with his top choice. Therefore, if no restrictions are placed on who can approach whom, then there is no unilaterally stable pairing. In this paper, we take the approach that stability in such situations can be restored by restricting each agent's ability to approach others. A familiar conservative norm forbids approaching any matched individual. Such a norm solves all instability problems, but at the cost of dramatically restricting individual liberty. We instead examine less restrictive norms that echo familiar social institutions of **taboos**, **status** and **power**.

We propose and analyze several solution concepts that employ these social institutions to restrict unilateral moves to obtain harmony. In the first, referred to as a T-equilibrium, agents adhere to social norms that determine which matches are permissible and which are taboo. The next two equilibrium notions are inspired by competitive equilibrium concepts, except that a status ranking prevails instead of prices. In detail, in a C-equilibrium, an agent can only approach agents who are lower-ranked than *himself*, while in an S-equilibrium, he can only approach agents who are lower-ranked than his *partner*. Finally, in the three J-equilibrium variants, in an equilibrium, a power hierarchy is determined which dictates who can approach whom. As a mnemonic device, we associate the letter T with taboo, S with status, C with competitive, and J with jungle.

As will be discussed later, the solution concepts we propose in the paper are related to some we have proposed in previous papers. In Richter and Rubinstein (2015, 2020) we have discussed the relationship between our solution concepts and competitive equilibrium in a market model. In Richter and Rubinstein (2021) we compared our approach to that of Nash equilibrium in a political situation which can be presented as a noncooperative game. In this paper, we apply, and sometimes adjust, the solution concepts to a family of cooperative games (matching problems) in which only exclusive coalitions of size two can be formed. The solution concepts allow us to formulate our primary innovation — harmony is disturbed by unilateral moves rather than by joint decisions (as in pairwise stability).

We do not deny that pairwise stability is often a natural solution concept. However, we also do not believe that it should be the only criterion for harmony that might be of interest in matching models.[1] We view our approach as complementary to pairwise stability. As a general philosophy regarding modeling in economics, we dispute the approach that there is one "right" solution concept. A modeler must ask himself what are the assumptions behind a solution concept before he applies it to his model. His choice of a solution concept should depend on the scenario he has in mind. In this paper, we propose and analyze several solution concepts based on unilateral stability, without ranking their importance in any way.

2. A leading example

Throughout, the analysis will be illustrated using a *common-ranking two-sided matching problem*. The two sides are N_1 and N_2. There are two common rankings, $i_1 \succ_1 i_2 \succ_1 \dots \succ_1 i_{n/2}$ over N_1 and $j_1 \succ_2 j_2 \succ_2 \dots \succ_2 j_{n/2}$ over N_2. Every agent in N_1 ranks his potential partners in N_2 by $\succ_2$ (that is, for each $i \in N_1$ and every $j,k \in N_2$, $j \succ^i k$ if and only if $j \succ_2 k$) and all of the agents in N_1 below them. Likewise, every agent in N_2 ranks his potential partners in N_1 by $\succ_1$ and all of the agents in N_2 below them.

A pairing is *mixed* if each agent is matched with an agent from the other side. In any two-sided matching problem, including this example, any Pareto optimal pairing must be mixed. Otherwise, there are two agents from one side who are matched and therefore there must also be two agents on the opposite side who are matched, and all four can be beneficially re-paired with members of the other group.

In the common-ranking two-sided matching problem every mixed pairing is Pareto optimal: if someone moves up in the common ranking of his partner, someone else must go down.

[1] Several papers consider modifications of pairwise stability where the voluntary formation of a new match is still the threat to stability (see for example: Morrill (2010), Root and Ahn (2020) and Tan (1990)).

3. The taboo equilibrium

The first solution concept, T-equilibrium, is inspired by Richter and Rubinstein (2020). It involves a social norm that determines which couples are permissible and which are taboo. A T-equilibrium is comprised of a pairing and a permissible set of matches, such that: (i) each agent's partner is his most preferred from among those with whom he can form a permissible pair, and (ii) the set of permissible matches is maximal in the sense that for any strict superset of permissible matches there is no pairing for which (i) is true.

If everything is permissible, then harmony is typically impossible. If a single pairing is permissible, then harmony is achieved, but at a terrible cost to freedom. The T-equilibrium concept strikes a balance: it allows as much freedom as possible without compromising the achievement of harmony.

Given a set of permissible matches Y, agent i's choice set is $C_i(Y) = \{j \mid i \leftrightarrow j \in Y\}$, which is the set of partners with whom he can form a permissible match. Notice that if i is permitted to choose j, then j is permitted to choose i. A *para-T-equilibrium* is a tuple $\langle Y, (x^i)\rangle$ where Y is a set of couples and (x^i) is a pairing such that for every agent i, agent x^i is i's most preferred in $C_i(Y)$. A *T-equilibrium* is a para-T-equilibrium such that there is no para-T-equilibrium $\langle Z, (y^i)\rangle$ with a strictly larger permissible set $Y \subset Z$.

Claim 1 characterizes the set of T-equilibrium pairings as the set of Pareto-optimal pairings (such pairings always exist).

Claim 1. *A pairing is a T-equilibrium pairing if and only if it is Pareto-optimal.*

Proof. Given a pairing (x^i), define $L((x^i))$ to be the set of all matches $i \leftrightarrow j$ such that $x^i \succsim^i j$ and $x^j \succsim^j i$. That is, it is the set of couples that are weakly less preferred by both of the involved agents relative to the pairing (x^i). Note that, for every i, the match $i \leftrightarrow x^i$ is in $L((x^i))$. Clearly, $\langle L((x^i)), (x^i)\rangle$ is a para-T-equilibrium since all pairs whose permissibility would disrupt the harmony of the pairing (x^i) are forbidden.

Let $\langle Y, (x^i)\rangle$ be a T-equilibrium. By the maximality of Y, it must be that $Y = L((x^i))$. If there exists a pairing (y^i) that Pareto-dominates (x^i), then the tuple $\langle L((y^i)), (y^i)\rangle$ is a para-T-equilibrium. Obviously, $L((y^i)) \supseteq L((x^i))$ and this inclusion is strict, since at least one agent, say j, is strictly better off in (y^i), which means that the match $j \leftrightarrow y^j$ is in $L((y^i))$ but not in $L((x^i))$. This establishes the existence of a para-T-equilibrium with a larger set of permissible pairs, contradicting $\langle Y, (x^i)\rangle$ being a T-equilibrium.

On the other hand, let (x^i) be a Pareto-optimal pairing. As previously mentioned, $\langle L((x^i)), (x^i)\rangle$ is a para-T-equilibrium. To demonstrate that it is a T-equilibrium, suppose to the contrary that there exists a para-T-equilibrium $\langle Y, (y^i)\rangle$ with $Y \supset L((x^i))$. In this scenario, all agents are weakly better off in (y^i) since for each agent i, agent x^i is available in Y. Moreover, the set Y contains at least one potential match $i \leftrightarrow j$ not in $L((x^i))$ for which $j \succ^i x^i$. However, in that case $y^i \succsim^i j \succ^i x^i$ and thus (y^i) Pareto-dominates (x^i), a contradiction. $\square$

The example. In the common-ranking two-sided matching problem, the Pareto-optimal pairings are precisely all mixed pairings. By the above claim the same is true regarding the T-equilibrium pairings. Some of the Pareto-optimal pairings require the permissible set to be very restrictive. In particular, the pairing $\{i_1 \leftrightarrow j_1, i_2 \leftrightarrow j_2, \dots\}$ requires the permissible set of the T-equilibrium to contain only the equilibrium matches.

4. Status ranking

This section examines scenarios in which harmony is achieved by means of a status ranking among the agents. Such rankings are prevalent in real life. Sometimes, they are external and due to factors outside of our domain, such as caste or physical appearance. Here, we have in mind situations in which the ranking is endogenous, just as prices are in a competitive equilibrium.

The following two subsections examine two solution concepts that utilize status as a means of achieving harmony. In the first, an agent can consider being matched only with agents of a weakly lower status than his own. In the second, an agent's "wealth" is determined in equilibrium by the status of his matched partner, and no agent can approach any other agent who has a higher status than his partner's.

In the spirit of competitive equilibrium, a status ranking can be thought of as a measure of value, and an agent chooses his optimal match given his own value (in the first solution concept) or that of his partner (in the second solution concept). In this interpretation, status imposes a physical restriction on an agent's choices.

An alternative interpretation of the ranking is that a *lower* rank corresponds to *more* prestige (under this interpretation, the ranking is the inverse of prestige, higher-ranked agents are less prestigious). An agent cannot bear the idea of being matched with anyone who is less prestigious than his current partner and therefore will only consider approaching agents who are weakly more prestigious, i.e. weakly lower-ranked, than his current partner. In this interpretation, status distorts an agent's view of potential replacements.

4.1. C-equilibrium

A C-equilibrium candidate is a tuple $\langle \trianglerighteq, (x^i)\rangle$ where $\trianglerighteq$ is an ordering over the set N (i.e. a complete, reflexive and transitive binary relation) and (x^i) is a pairing. The statement $i \trianglerighteq j$ is interpreted as "i has a weakly higher status than j". A social institution prevents an agent from being matched with anyone who is strictly higher ranked than *himself*.

In a *C-equilibrium*, for every agent i, his partner x^i is i's most preferred partner in his choice set $\{j \in N - \{i\} \mid i \trianglerighteq j\}$. Thus, the ordering $\trianglerighteq$ determines each agent's "budget set". In order to characterize C-equilibria, we introduce the concept of *pair-rankability*. A matching problem is said to be *pair-rankable* if there exists a partition of N into doubletons $\{I_1, \dots, I_{n/2}\}$ with the property that, for every i and q, if $i \in I_q$, then i top-ranks his doubleton's partner from the union of the later doubletons $I_q \cup \dots \cup I_{n/2}$.[2] If a matching problem is pair-rankable, then the pairing generated by such a partition is clearly the unique pairwise stable pairing.

The following claim demonstrates that a C-equilibrium exists if and only if the matching problem is pair-rankable.

Claim 2. *(i) A matching problem has a C-equilibrium if and only if it is pair-rankable.*

(ii) If a C-equilibrium pairing exists, then it is unique and pairwise stable (the ordering which supports the pairing need not be unique).

Proof. (i) Let $\langle \trianglerighteq, (x^i) \rangle$ be a C-equilibrium and let i be a $\trianglerighteq$-maximal agent. Agents i and x^i must be equally top $\trianglerighteq$-ranked, and thus both agents must top-rank each other. These two agents form I_1 and are removed. This process repeats itself with the remaining agents to eventually form $I_2, \dots, I_{n/2}$. Thus, the matching problem is pair-rankable.

Assume that the matching problem is pair-rankable with a sequence of doubletons $I_1, \dots, I_{n/2}$. Define $i \trianglerighteq j$ if $i \in I_q$, $j \in I_r$ and $q \le r$. Define $x^i = j$ if $\{i, j\}$ is one of the doubletons. Clearly, the tuple $\langle \trianglerighteq, (x^i) \rangle$ constitutes a C-equilibrium and its pairing is pairwise stable.

(ii) A proof by induction on the number of agents: Assume that a C-equilibrium exists. As in (i), the highest-status agent is matched with another highest-status agent, and they both must top-rank each other (recall that preferences are strict). Consequently, they must be matched in any C-equilibrium (since for any equilibrium ranking, one of the two agents can "afford" the other and would prefer him over any other potential partner). Any C-equilibrium induces a C-equilibrium among the remaining agents. By the induction hypothesis, this induced C-equilibrium pairing is unique among the remaining agents. Therefore, the C-equilibrium pairing is unique, and by part (i) it is pairwise stable. However, the ranking need not be unique. $\square$

The example. The common-ranking two-sided matching problem is uniquely pair-rankable with $I_q = \{i_q, j_q\}$. By Claim 2, there is a unique C-equilibrium: the pairing is $\{i_1 \leftrightarrow j_1, i_2 \leftrightarrow j_2, \dots\}$ and the status ranking is $i_1 \sim j_1 \triangleright i_2 \sim j_2 \triangleright \dots \triangleright i_{n/2} \sim j_{n/2}$.

4.2. S-equilibrium

An S-equilibrium candidate also consists of an ordering of the agents and a pairing. As before, the ordering is interpreted as reflecting status, but now the social institution prevents an agent from approaching anyone who is ranked higher than *his partner*. In an equilibrium, no agent can find a different partner who he judges to be more desirable and who has a weakly lower status than his current partner. Formally, an *S-equilibrium* is a tuple $\langle \trianglerighteq, (x^i) \rangle$ where $\trianglerighteq$ is an ordering over N and (x^i) is a pairing such that for every agent i, there is no j such that $j \succ^i x^i$ and $x^i \trianglerighteq j$. Note that every C-equilibrium is an S-equilibrium.

In this context, we find the previously mentioned anti-prestige interpretation particularly compelling: an agent i will approach j only if he prefers j to his current partner ($j \succ^i x^i$) and j is weakly more prestigious than his current partner ($x^i \trianglerighteq j$).

The S-equilibrium concept is closely related to the abstract equilibrium concept discussed in Richter and Rubinstein (2015). However, there it was assumed that the set of feasible profiles is closed under all permutations, which never holds for matching problems. Therefore, Claim 3(i) below does not follow from our work there.

A different interpretation of the S-equilibrium notion views a match as a kind of double ownership. When agents A and B are matched, A owns B, and simultaneously, B owns A! Agents are ranked according to some notion of value whereby each agent "owns" his partner, and can "exchange" him for any weakly "less expensive" agent. In an S-equilibrium, no agent wishes to do so. Double ownership might seem strange at first glance, but consider a street of identical duplexes where each resident owns one unit. If all units are the same, then each duplex can be viewed as a partnership, with the only distinction between units being who your neighbor is. Selling a unit means selling the right (or duty) to be someone's neighbor. In this respect, two individuals living in the same duplex are involved in double ownership.

We now show that every S-equilibrium pairing is Pareto-optimal and that the S-equilibrium notion is quite different from pairwise stability. Even the existence of one concept does not imply the existence of the other:

Claim 3. *(i) Every S-equilibrium pairing is Pareto-optimal.*

(ii) There is a matching problem with an S-equilibrium but no pairwise-stable pairing.

(iii) There is a matching problem with a pairwise-stable pairing but no S-equilibrium.

Proof. (i) Recall that all preferences are assumed to be strict. Let $\langle \trianglerighteq, (x^i) \rangle$ be an S-equilibrium, and (y^i) be a pairing such that $y^i \succsim^i x^i$ for all i. Let i_1 be an agent with the highest $\trianglerighteq$-ranked partner in the pairing (x^i). Agent i_1 is the "wealthiest" and therefore

[2] Pair-rankability is a milder condition than α-reducibility (Alcalde, 1995), which is used to ensure the existence of a pairwise-stable pairing in the roommate problem.

his partner x^{i_1} is i_1's first-best and $y^{i_1} = x^{i_1}$. Let i_2 be an agent with the highest $\trianglerighteq$-ranked partner among $N - \{i_1, x^{i_1}\}$. Again, it must be that $y^{i_2} = x^{i_2}$. Repeating this argument $n/2$ times leads to the conclusion that $(y^i) = (x^i)$.

(ii) Let $N = \{1, 2, 3, 4\}$. The following is Gale and Shapley (1962)'s canonical example of a roommate problem without a pairwise-stable pairing[3]:

Table 1
A matching problem with an S-equilibrium but no pairwise-stable pairing.

Agent	1	2	3	4
1st Preference	2	3	1	1
2nd Preference	3	1	2	2
3rd Preference	4	4	4	3

There are two S-equilibrium pairings: $\{1 \leftrightarrow 4, 2 \leftrightarrow 3\}$ (see Table 1) and $\{1 \leftrightarrow 3, 2 \leftrightarrow 4\}$. Both are supported by the ordering $1 \rhd 2 \rhd 3 \rhd 4$. The other pairing $\{1 \leftrightarrow 2, 3 \leftrightarrow 4\}$ is not an S-equilibrium pairing under any ordering $\trianglerighteq$ since it must be that $3 \rhd 1$ (to prevent 2 from approaching 3), and it must be that $1 \rhd 3$ (to prevent 4 from approaching 1).

(iii) Consider the following matching problem with $N = \{1, 2, 3, 4\}$:

Table 2
A matching problem with a pairwise stable pairing but no S-equilibrium.

Agent	1	2	3	4
1st Preference	2	3	4	1
2nd Preference	3	4	1	2
3rd Preference	4	1	2	3

The pairing $\{1 \leftrightarrow 3, 2 \leftrightarrow 4\}$ is the unique pairwise-stable pairing (see Table 2). No S-equilibrium exists: given any ranking $\trianglerighteq$, one of the agents, and without loss of generality let it be 1, is matched with his first-best and the resulting pairing is $\{1 \leftrightarrow 2, 3 \leftrightarrow 4\}$. However, it must then be that both $3 \rhd 1$ (to prevent 2 from approaching 3) and $1 \rhd 3$ (to prevent 4 from approaching 1). □

The example. By Claim 3, an S-equilibrium pairing is Pareto optimal, namely mixed. On the other hand, any mixed pairing is a S-equilibrium pairing. In particular, the pairing $\{i_1 \leftrightarrow j_{n/2}, i_2 \leftrightarrow j_{n/2-1}, \dots\}$ with the ranking $i_1 \rhd i_2 \rhd \cdots \rhd i_{n/2} \rhd j_1 \rhd j_2 \rhd \cdots \rhd j_{n/2}$ is an S-equilibrium in which the least-popular agent in one group, $j_{n/2}$, is the "wealthiest" since he has the most prestigious partner. This contrasts with the C-equilibrium in which there is a unique equilibrium pairing and wealth and popularity are fully aligned.

5. Jungle equilibria

Besides taboos and status, another force that governs societies is power (see Piccione and Rubinstein (2007)'s discussion of the Jungle Economy). By "power", we are not referring exclusively to raw physical strength, but also to gentler and subtler forms, such as seniority, conversational ability and charm. Power is modeled here as a strict ordering $\rhd$ over the agents (that is, a transitive and antisymmetric binary relation) where $a \rhd b$ means that a is more powerful than b. This section presents three jungle equilibrium variants. In each of them, an equilibrium candidate is a tuple $\langle \rhd, (x^i) \rangle$ where $\rhd$ is a strict ordering of N and (x^i) is a pairing.

Power limits the ability of agents to approach one another. The most obvious limitation is that a weaker agent cannot approach a stronger one, which is the basis of the J1-equilibrium notion. But power can be more intricate. In the J2-equilibrium, an agent is only able to approach another if both the approached agent and the approached agent's partner are weaker than him. The J3-equilibrium takes this a step further, by adding that the approaching agent also has to be stronger than his current partner whom he seeks to abandon.

The assumption that the power relation is a *strict* ordering is important. If the power relation is only required to be an ordering, then "anything goes": any pairing together with the total indifference relation would be a Jungle equilibrium for all three variants because the indifference between every two agents prevents any agent from using power.

5.1. J1-equilibrium

A *J1-equilibrium* is a tuple $\langle \rhd, (x^i) \rangle$ for which there are no two agents, i and j, such that i is stronger than j and i strictly prefers j over his current partner x^i. That is, an agent is deterred from approaching another not by a fear of rejection (as in the case of pairwise stability), but rather by the power of the desired agent.

[3] Consider any pairing. Let i be the agent matched with 4. Agent i prefers every other agent to 4 and there is $j \in \{1, 2, 3\}$ who top-ranks i. Thus, the pair (i, j) blocks the profile from being pairwise-stable.

The J1-equilibrium concept is weaker than the C-equilibrium concept: if $\langle \trianglerighteq, (x^i) \rangle$ is a C-equilibrium, then $\langle \blacktriangleright, (x^i) \rangle$ is a J1-equilibrium, where $\blacktriangleright$ is any strict tie-breaking of $\trianglerighteq$. We now show that any J1-equilibrium outcome is pairwise stable (and consequently, Pareto optimal). However, the J1-equilibrium concept is more stringent, and may fail to exist even when a pairwise-stable pairing exists.

Claim 4. *(i) Every J1-equilibrium pairing is pairwise stable.*

(ii) A J1-equilibrium need not exist, even when a pairwise-stable pairing exists.

Proof. (i) Let $\langle \triangleright, (x^i) \rangle$ be a J1-equilibrium. Suppose that there are two agents i and j who strictly prefer each other to their current partners. Then the stronger agent prefers the weaker one over his current partner, violating the J1-equilibrium condition.
(ii) In the example from Table 2, the unique pairwise-stable pairing is $\{1 \leftrightarrow 3, 2 \leftrightarrow 4\}$ and by part (i) this is the only J1-equilibrium candidate. However, it is not a J1-equilibrium pairing since if it were, then the strongest agent would be paired with his first-best and no agent is. □

The example. In the common-ranking two-sided matching problem, there is a unique J1-equilibrium pairing $\{i_1 \leftrightarrow j_1, i_2 \leftrightarrow j_2, \cdots, i_{n/2} \leftrightarrow j_{n/2}\}$ (supported by any power ordering satisfying $i_1, j_1 \triangleright i_2, j_2 \triangleright \cdots \triangleright i_{n/2}, j_{n/2}$). Here's why: Since every agent in N_2 desires i_1, he must be stronger than everyone in N_2 (except perhaps his partner), which means that i_1 has to be matched with his first-best, namely j_1. Similarly, j_1 must be stronger than all members of N_1 except possibly i_1. This pattern continues down the ranking. Among the remaining agents, i_2 and j_2 are matched and i_2 must be more powerful than $\{j_3, \dots, j_{n/2}\}$ while j_2 must be stronger than $\{i_3, \dots, i_{n/2}\}$ and so on.

5.2. The adapted serial dictatorship algorithm

The *adapted serial dictatorship algorithm* defines a pairing for any arbitrary priority ordering over the agents. The algorithm proceeds as follows: the top priority agent a_1 is paired with his first-best, b_1, and both are removed. Next, from among those that haven't been paired, the top priority agent a_2 is paired with his first-best, b_2, and both are removed. Continue in this manner to obtain a pairing $\{a_1 \leftrightarrow b_1, \dots, a_{n/2} \leftrightarrow b_{n/2}\}$.

The algorithm is adapted in the sense that rather than choosing over objects or bundles, the objects of choice are the agents themselves, and if an agent is chosen, then he loses his ability to choose (whether or not he would choose his current partner). That is, half of the agents "make a choice" and the other half of the agents "are chosen". Note that the outcome of the algorithm is always Pareto optimal.

5.3. J2-equilibrium

A *J2-equilibrium* is a tuple $\langle \triangleright, (x^i) \rangle$ in which there are no two agents i and j such that i desires j more than his current partner ($j \succ^i x^i$) and is more powerful than both j and j's partner ($i \triangleright j$ and $i \triangleright x^j$). Now an agent is deterred not only by the strength of the approached agent, but also by the strength of that agent's partner. Obviously, every J1-equilibrium is also a J2-equilibrium.

Claim 5. *(i) A J2-equilibrium always exists.*

(ii) Every J2-equilibrium pairing is Pareto-optimal.

(iii) There can be a Pareto-optimal pairing that is not a J2-equilibrium pairing.

(iv) Every S-equilibrium pairing is a J2-equilibrium pairing.

Proof. (i) For an arbitrary priority ordering, run the adapted serial dictatorship algorithm to obtain $\{a_1 \leftrightarrow b_1, \dots, a_{n/2} \leftrightarrow b_{n/2}\}$. Define the power ordering to be $a_1 \triangleright a_2 \triangleright \cdots \triangleright b_2 \triangleright b_1$. Note that every a_k is stronger than every b_l.

This achieves a J2-equilibrium: For every l, agent b_l cannot approach any other agent because every couple includes an a_k stronger than b_l. For every k, if a_k prefers j over his partner, then j must have been removed in the above construction before a_k, and therefore, either j or j's partner is stronger than a_k.

(ii) Let $\langle \triangleright, (x^i) \rangle$ be a J2-equilibrium. Assume that (y^i) Pareto-dominates (x^i). Let j be the strongest agent in $M = \{i : x^i \neq y^i\}$. Then $y^j \succ^j x^j$ (recall that preferences are strict), and j is stronger than both y^j and y^j's original partner, x^{y^j} (because both are in M), violating $\langle \triangleright, (x^i) \rangle$ being a J2-equilibrium.

(iii) Consider again the matching problem described in Table 2. The pairing $\{1 \leftrightarrow 3, 2 \leftrightarrow 4\}$ is Pareto-optimal. If it were a J2-equilibrium outcome, then the strongest agent would be matched to his first-best, but no agent is matched with his first-best.

(iv) Let $\langle \trianglerighteq, (x^i) \rangle$ be an S-equilibrium and break ties so that $\trianglerighteq$ is strict. Define a power ranking $\blacktriangleright$ by ranking agents by the status of their partners: $i \blacktriangleright j$ if $x^i \triangleright x^j$. Suppose that for some i and j, it holds that $j \succ^i x^i$. Then, it must be that $j \triangleright x^i$ since $\langle \trianglerighteq, (x^i) \rangle$ is an S-equilibrium. But then $x^j \blacktriangleright i$. Therefore, $\langle \blacktriangleright, (x^i) \rangle$ is a J2-equilibrium.[4] □

The example. Recall that the set of the S-equilibrium pairings is the set of all mixed pairings. By Claim 5(ii, iv), this is also the set of J2-equilibrium pairings. This is in contrast to the uniqueness of the J1-equilibrium pairing.

On the other hand, not every power relation is a part of some J2-equilibrium. For example, when $n = 4$, there is no J2-equilibrium with the power relation $j_2 \triangleright i_1 \triangleright i_2 \triangleright j_1$. This is because j_2 is the strongest agent, and must be matched with i_1. Thus, the only candidate pairing is $\{i_1 \leftrightarrow j_2, i_2 \leftrightarrow j_1\}$ which is not a J2-equilibrium since i_1 prefers j_1, and is stronger than both i_2 and j_1.

5.4. J3-equilibrium

A *J3-equilibrium* is a tuple $\langle \triangleright, (x^i) \rangle$ in which there are no two agents i and j such that i prefers j over his current partner ($j \succ^i x^i$) and is more powerful than j, j's partner and his own partner ($i \triangleright j, x^i, x^j$). That is, for agent i to approach another agent j, he now also needs to be strong enough to leave his current partner.

Clearly, every J2-equilibrium is a J3-equilibrium. By Claim 6 below, the set of J3-equilibrium pairings is identical to those of the J2-equilibrium, but unlike the case of J2-equilibrium, every power relation is part of some J3-equilibrium.

Claim 6. *(i) For every power relation $\triangleright$, there is a unique J3-equilibrium $\langle \triangleright, (x^i) \rangle$.*

(ii) The set of J3-equilibrium pairings is identical to the set of J2-equilibrium pairings.

Proof. (i) Fix a power relation $\triangleright$. Run the adapted serial dictatorship algorithm with $\triangleright$ being the priority ordering. To see that the obtained pairing, together with $\triangleright$, is a J3-equilibrium, notice that any agent who "makes a choice" can only prefer earlier removed agents, i.e. those who are stronger than him or are paired with a stronger partner than him. Any agent who "is chosen" is incapable of approaching any other agent because he is matched with a stronger agent.

For uniqueness, suppose there are two different J3-equilibria with the same power relation, $\langle \triangleright, (x^i) \rangle$ and $\langle \triangleright, (y^i) \rangle$. Let j be the $\triangleright$-strongest agent in $M = \{i : x^i \neq y^i\}$. Without loss of generality, suppose that $x^j \succ^j y^j$. The tuple $\langle \triangleright, (y^i) \rangle$ is not a J3-equilibrium because j prefers x^j to y^j, and $j \triangleright y^j, x^j, y^{x^j}$ since y^j, x^j and y^{x^j} are all in M.

(ii) As mentioned, every J2-equilibrium is also a J3-equilibrium. To see the reverse, by the construction and uniqueness of part (i), every J3-equilibrium pairing is constructable by running the adapted serial dictatorship. From Claim 5, running the adapted serial dictatorship always results in a J2-equilibrium pairing (potentially with a different power relation). Therefore, every J3-equilibrium pairing is a J2-equilibrium pairing. □

5.5. External vs. internal power relations

In the original jungle equilibrium concept of Piccione and Rubinstein (2007), the power relation between agents is external. In contrast, in our definitions of jungle equilibria (and as in Rubinstein and Yıldız (2022)), power is determined internally as part of the equilibrium concept. This is akin to the endogeneity of prices in the Walrasian setting and the endogeneity of the permissible set and status ranking in our other solution concepts.

Note that not all power relations are consistent with the J1- or J2-equilibrium concepts, just as not all price vectors constitute a Walrasian equilibrium. On the other hand, regarding the J3-equilibrium concept, Claim 6 shows that any power relation corresponds to a unique equilibrium outcome, and therefore this solution concept fits both internal and external power structures.

6. Final comments

As mentioned in the introduction, we do not argue that any of the proposed solution concepts is superior or inferior to pairwise stability, on either normative or positive grounds. We have intentionally avoided any speculation as to which of the examined institutions might be more likely to emerge. Rather, the proposed solution concepts express fundamentally different concerns for harmony in a society, with each concept narrating its own story. We conclude with three comments:

6.1. Two types of equilibria

The equilibrium concepts we have presented fall into two categories: those resembling market equilibrium where agents face "budget sets" and the optimal *choice* profile must be feasible, and those that are more akin to non-cooperative game-theoretical solution concepts where an equilibrium is a feasible profile that is immune to certain classes of *unilateral deviations*. Let us elaborate.

[4] The J2-equilibrium notion requires that i be stronger than both the agent he approaches and that agent's partner. This argument shows that if we define an alternate J1-equilibrium notion using only the latter condition then this alternate notion is equivalent to an S-equilibrium.

In a standard competitive equilibrium, every agent possesses an endowment, prices are endogenously determined, agents make optimal choices based on their endowment and the prices without regard to the choices made by others, and their choices are coherent. Our T-, C-, and J1-equilibrium concepts align with this framework. In these concepts, an additional price-like element – either the set of permissible alternatives, status ranking, or power ranking – is formed endogenously. This element dictates what options agents can choose (their "budget sets") and the agents select from their options according to their preferences, without regard to the choices made by others. In an equilibrium, the profile of choices is feasible.

On the other hand, the S-, J2-, and J3-equilibrium concepts are predicated on immunity to deviations (as in non-cooperative game theory). In these concepts, a candidate pairing constitutes an equilibrium if it is immune to a class of unilateral deviations, as determined by an endogenous price-like element (either a status or power ranking). For these concepts (and pairwise stability as well), the set of deviations that agents can make depends upon the equilibrium configuration.

6.2. Algorithms

A major attraction of pairwise stability in the two-population matching problem is that Gale and Shapley (1962)'s "deferred acceptance" algorithm achieves either the male- or female-optimal pairwise-stable pairing. However, this algorithm only works for the two-population matching problem.

We have proposed constructive algorithms for the C-, J2-, and J3-equilibrium concepts. We find the adapted serial dictatorship algorithm (used in Claims 5 and 6 to construct an equilibrium pairing) to have interesting features: (i) some agents are dictators while others are dictated to and (ii) any J2/J3-equilibrium pairing can be obtained in this way by adjusting the parameters of the algorithm, that is, the priority ordering.

6.3. Relationship between the solution concepts' outcomes

Fig. 1 illustrates the connections between the various equilibrium concepts introduced in this paper and how they relate to Pareto-optimality and pairwise stability.[5]

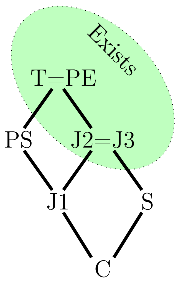

Fig. 1. Weak inclusion relationship between the equilibrium notions, pairwise stability (PS) and Pareto efficiency (PE). J2-, J3- and T-equilibria always exist.

Existence always holds for the J2-, J3- and T-equilibrium concepts. The First Welfare Theorem (all equilibrium pairings are Pareto optimal) holds for all of the solution concepts. The Second Welfare Theorem (all Pareto-optimal pairings are equilibrium pairings) is guaranteed only for T-equilibrium.

Declaration of competing interest

The author has no relevant source of financial support to disclose.

IRB approval was not obtained as no experiments were conducted and no human subject data was used.

Data availability

No data was used for the research described in the article.

References

Alcalde, José, 1995. Exchange-proofness or divorce-proofness? Stability in one-sided matching markets. Econ. Des. 1, 275–287.
Gale, David, Shapley, Lloyd, 1962. College admissions and the stability of marriage. Am. Math. Mon. 69, 9–15.
Morrill, Thayer, 2010. The roommates problem revisited. J. Econ. Theory 145, 1739–1756.

[5] All of the depicted relationships have been demonstrated in the previous claims except for the non-inclusion of the S- and J1-equilibrium notions. Table 2 presents a case with an S-equilibrium but no J1-equilibrium. It is straightforward to formulate an example of the opposite situation. In general, all inclusions are weak, however for each inclusion, it is simple to come up with an example in which the inclusion is strict.

Piccione, Michele, Rubinstein, Ariel, 2007. Equilibrium in the jungle. Econ. J. 117, 883–896.
Richter, Michael, Rubinstein, Ariel, 2015. Back to fundamentals: equilibrium in abstract economies. Am. Econ. Rev. 105, 2570–2594.
Richter, Michael, Rubinstein, Ariel, 2020. The permissible and the forbidden. J. Econ. Theory 188, 105042.
Root, Joseph, Ahn, David S., 2020. Incentives and Efficiency in Constrained Allocation Mechanisms. mimeo.
Rubinstein, Ariel, Yıldız, Kemal, 2022. Equilibrium in a civilized jungle. Theor. Econ. 17, 943–953.
Tan, Jimmy J.M., 1990. A maximum stable matching for the roommates problem. BIT Numer. Math. 30, 631–640.