Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Mechanism Design for Team Formation

Mason Wright
Computer Science & Engineering
University of Michigan
Ann Arbor, MI
masondw@umich.edu

Yevgeniy Vorobeychik
Electrical Engineering and Computer Science
Vanderbilt University
Nashville, TN
yevgeniy.vorobeychik@vanderbilt.edu

Abstract

Team formation is a core problem in AI. Remarkably, little prior work has addressed the problem of mechanism design for team formation, accounting for the need to elicit agents' preferences over potential teammates. Coalition formation in the related hedonic games has received much attention, but only from the perspective of coalition stability, with little emphasis on the mechanism design objectives of true preference elicitation, social welfare, and equity. We present the first formal mechanism design framework for team formation, building on recent combinatorial matching market design literature. We exhibit four mechanisms for this problem, two novel, two simple extensions of known mechanisms from other domains. Two of these (one new, one known) have desirable theoretical properties. However, we use extensive experiments to show our second novel mechanism, despite having no theoretical guarantees, empirically achieves good incentive compatibility, welfare, and fairness.

Introduction

Teamwork has been an important and often-studied area of artificial intelligence research. Typically, the focus is on coordinating agents to achieve a common goal. The complementary problem of team formation considers how to form high-quality teams, whose agents have skills that are jointly well suited for a task (Marcolino, Jiang, and Tambe 2013). Notable team formation applications include formation of research teams, class project groups, groups of roommates, or disaster relief teams.

Many prior team formation studies have assumed that agents are indifferent about which other agents they are teamed with, or have preferences known to the team formation mechanism. Models dealing with known agent preferences over teammates, termed *hedonic games*, have seen an extensive literature since being introduced by Aumann and Dreze (1974). In a hedonic game, the mechanism is given a set of agents, each having public preferences over which others might be on its team; the mechanism must partition the agents into teams based on their preferences.

Past research on hedonic games has focused on the problem of forming *stable* coalitions, from which no set of agents would prefer to defect. Since a core partition may not exist in a hedonic game, even when preferences of players are additively separable (Banerjee, Konishi, and Sönmez 2001), much research is focused on alternative notions of stability, or on highly restricted agent preferences (Bogomolnaia and Jackson 2002; Alcalde and Revilla 2004; Cechlarova and Romero-Medina 2001), or on the time complexity of testing core emptiness (Ballester 2004; Sung and Dimitrov 2010).

We consider team formation as a mechanism design problem, where individuals have preferences over teammates, as in hedonic games. As in traditional mechanism design (and unlike hedonic games), we assume that these preferences are private and must be elicited in order to partition players reasonably into teams. We draw a connection to another budding literature, that of combinatorial matching market design, which has course allocation as a typical application (Budish and Cantillon 2012).

An important concern in combinatorial matching, which we inherit, is the ex post fairness of allocations. For example, consider a simple randomized mechanism, *random serial dictatorship*, which has been proposed for course allocation and is readily adapted to team formation. In random serial dictatorship, agents are randomly ordered by the mechanism and then take turns, in order, selecting their entire teams from among the remaining agents. Random serial dictatorship is *strategyproof*, meaning that it is a dominant strategy for any agent to report its true preferences over teams. Random serial dictatorship is also *ex post Pareto efficient*, in that any allocation it returns cannot be modified to improve an agent's welfare without reducing some other agent's (assuming no indifferences). But this mechanism results in a highly inequitable distribution of outcomes ex post.

Budish and Cantillon (2012) proposed a more sophisticated alternative, *approximate competitive equilibrium from equal incomes (A-CEEI)*, which is *strategyproof-in-the-large* (i.e., when the number of players becomes infinite), and provably approximately fair (Budish and Cantillon 2012; Budish 2011). The work on combinatorial matching in turn follows earlier work on bipartite matching and school choice (Roth and Peranson 1999; Abdulkadiroglu and Sönmez 2003).

Our contributions are as follows.

1. We present the problem of mechanism design for team formation, focused on achieving (near-)incentive compatible preference reporting, high social welfare, and fair allocation. This problem is closely related to both combinatorial and bipartite matching market design, but is distinct from both in two senses: first, the matching is not bipartite (players match to other players), and therefore typical matching algorithms which only guarantee strategyproofness for one side are unsatisfactory; and second, mechanisms used in combinatorial exchanges to provide fairness guarantees are not directly applicable, as they rely on having a fixed set of items which are the subject of the match and which are not themselves strategic;
2. we extend two well-known mechanisms (random serial dictatorship and Harvard Business School draft) used for combinatorial matching to our setting;
3. we propose two novel mechanisms for our setting (A-CEEI for team formation, or A-CEEI-TF, and one-player-one-pick draft, or OPOP);
4. we prove that A-CEEI-TF is approximately fair and strategyproof-in-the-large;
5. we offer empirical analysis of all mechanisms, which shows that our second mechanism, OPOP, outperforms others on most metrics, and has better incentive properties than A-CEEI-TF.

An important and surprising finding of our investigation is that the simple draft mechanism we propose empirically outperforms the more complex A-CEEI-TF alternative by a large margin in fairness and incentive compatibility, even while A-CEEI-TF has more compelling theoretical guarantees.

Mechanism Design Problem

Our point of departure is the formalism of *hedonic games*. We define a *hedonic game* as a tuple $(N, \succ)$, where N is the set of players, and $\succ$ is a vector containing each player's preference order over sets of other players that it could be teamed with. The task is to partition the players in N into a coalition structure, where each player is in exactly one coalition.

We assume that player preferences are *additively separable* (Aziz, Brandt, and Seedig 2011), which means that there exists an assignment of values $u_i(j)$ for all players i and their potential teammates j, so that i's total utility of a subset of others S is $\sum_{j \in S} u_i(j)$ (which induces a corresponding preference ordering over subsets of possible teammates). In addition, we assume that $u_i(j) \geq 0$ for all i, j.

These assumptions are useful for two reasons. First, in many data sets that record preferences of individuals over others, the preferences are entered as non-negative values for individuals, as in rank order lists or Likert ratings of individuals. Additive separable preferences are the most natural way to induce preferences over groups from such data. Second, many prior studies in team formation and the related domain of course allocation have assumed that agents have non-negative, additive separable preferences, as in the $\mathcal{B}$-preferences of (Cechlarova and Romero-Medina 2001) and in the bidding points auction.

Most prior work on hedonic games focuses on coalition stability. Our goal is distinct: We take as input player preferences over teams (that is, over others that they could be teamed with), which we assume to be additive with non-negative values, and output a partition of the players into teams. We assume that it is subsequently difficult for players to alter team membership. Our primary challenge, therefore, is to encourage players to report their preferences honestly, and form teams that are fair and yield good teammate matchings; all three notions shall be made precise presently. Note that in this construction we assume that no money can change hands (unlike the work by Li et al. (2004)).

Observe that in our model, all players always prefer to be put on a single team (since values for all potential teammates are positive). In reality, many team formation problems have hard constraints on team sizes (or, equivalently, on the number of teams), particularly when multiple tasks need to be accomplished. For example, project teams usually have an upper bound on size. We capture this by introducing team size constraints; formally, the size of any team must be in the interval $[\underline{k}, \overline{k}]$, with $\underline{k} \geq 1$, $\overline{k} \leq |N|$, and $\underline{k} \leq \overline{k}$. For example, if a classroom with 25 students must be divided into 6 approximately equal-size teams, we could have $\underline{k} = 4$ and $\overline{k} = 5$. We assume throughout that the specific values of $\underline{k}$ and $\overline{k}$ admit a feasible allocation. (This is not always the case; see supplemental material for details.)

In contrast with a typical approach in mechanism design, which seeks to maximize a single objective such as social welfare or designer revenue, subject to a constraint set, we take an approach from the matching market design literature, and seek a collection of *desirable properties* (see, e.g., Budish (2012)). Specifically, we consider three properties: incentive compatibility, social welfare, and fairness. Given the fact that all three cannot be achieved simultaneously in our setting, we will analyze the extent to which each can be achieved through specific mechanisms.

Incentive Compatibility Incentive compatibility holds if there is no incentive for an agent to misreport its preferences. We consider two forms of incentive compatibility: *strategyproofness*, which means that it is a dominant strategy for any agent to report its true preferences, and *ex post equilibrium*, which means that it is a Nash equilibrium for all agents to report their true preferences. The former will be considered in theoretical analysis, while the latter will be the focus of empirical incentive assessment. In particular, our theory will focus on *strategyproofness-in-the-large* (Budish 2011), defined as follows. Consider a market where each agent has been replaced with a measure-one continuum of replicas of itself, such that each individual agent has zero measure and all agents are price takers. A mechanism is strategyproof-in-the-large if, in such a market, it is a dominant strategy for each agent to reveal its true preferences. An example of a mechanism that is not strategyproof-in-the-large is the Harvard Business School draft considered below, in which an

agent may benefit from misreporting its preferences, regardless of its own measure relative to the market size (Budish and Cantillon 2012). In empirical analysis, in contrast, we determine a lower bound on the *regret of truthful reporting*, which is the most any agent can gain ex post by misreporting preferences when all others are truthful.

Social Welfare As in traditional mechanism design, we consider social welfare as one of our primary design criteria. Social welfare is just the sum of player utilities achieved by a specific partition of players into teams. Formally, if $\mathcal{Q}$ is a partition of players, social welfare is defined as $SW(\mathcal{Q}) = \frac{1}{|N|}\sum_{S\in\mathcal{Q}}\sum_{i,j\in S} u_i(j)$. In addition, we consider the weaker notion of *ex post Pareto optimality* when discussing alternative mechanisms and their theoretical properties. A partition of players $\mathcal{Q}$ is ex post Pareto optimal if no other partition strictly improves some agent's utility without lowering the utility of any other agent.

Fairness The measure of fairness we consider is *envy-freeness*. An allocation is *envy-free* if each agent weakly prefers its own allocation to that of any other agent. An approximate notion of envy-freeness that we adopt from Budish (2011) is *envy bounded by a single teammate*, in which any allocation an agent prefers to their own ceases to be preferred through removal of a single teammate from it.[1] The following negative result makes apparent the considerable challenge associated with the design problem we pose.

Proposition 1. *There may not exist a partition of players that bounds envy by a single teammate.*

Proof Consider a team formation problem with 6 agents, $\{A, B, C, D, E, F\}$, $\underline{k} = 3$, $\overline{k} = 3$, so that two equal-size teams must be formed. The agents' additive separable preferences are encoded in Table 1.

	A	B	C	D	E	F
A	x	0	1	2	4	8
B	8	x	4	2	1	0
C	8	0	x	4	2	1
D	8	1	0	x	4	2
E	8	2	1	0	x	4
F	8	4	2	1	0	x

Table 1: Each row i encodes the additive separable value for agent i of each other agent.

No partition of these agents into two teams of size 3 gives every agent envy bounded by a single teammate. To see this, consider that each agent other than A has a bliss point on a team with A and one other agent, where the second agent is C for agent B, D for agent C, and so on until "wrapping around" with B for agent F. Three of the agents will not be on a team with agent A, and at least one of these agents, say agent i, will not be on a team with its second-favorite agent either. Some other agent j must then be on a team with the two most-preferred agents of the player i. By construction, player i is on a team of value 3 or less, while the team of agent j has value 12 to agent i, and value 4 to agent i with its more valuable player (player A) removed. Therefore, envy cannot be bounded by a single teammate for all agents. □

[1] In the supplemental material we discuss another measure of fairness.

Team Formation Mechanisms

We describe four mechanisms for team formation: two are straightforward applications of known mechanisms, while two are novel.

Random Serial Dictatorship

Random serial dictatorship (RSD) has previously been proposed in association with school choice problems (Abdulkadiroglu and Sönmez 2003). In RSD, players are randomly ordered, and each player chosen in this order selects his team (with players thereby chosen dropping out from the order). The process is repeated until all players are teamed up.

Proposition 2. *Random serial dictatorship is strategyproof, and ex post Pareto efficient as long as players choosing later cannot choose a larger team.* [2]

While RSD is ex post Pareto efficient, this turns out to be a weak guarantee, and does not in general imply social welfare maximization, something that becomes immediately apparent in the experiments below. Envy-freeness is, of course, out of the question due to Proposition 1.

Harvard Business School (HBS) Draft

Players are randomly ordered, with the first T assigned as captains. We then iterate over captains, first in the random order, then in reverse, alternating. The current team captain selects its most-preferred remaining player to join its team, based on its reported preferences.

Proposition 3. *HBS draft is not strategyproof or ex post Pareto efficient.*

One-Player-One-Pick (OPOP) Draft

Players are randomly ordered. Given the list of team sizes, the first T players are assigned to be captains of the respective teams. Then iterate over the complete player list. If the next player is a team captain, it selects its favorite unassigned agent to join its team. If the next player is unassigned, it will be assigned to join its favorite incomplete team (as defined below), and if the team still has space, this player chooses its favorite unassigned agent to join them. We define a "favorite" incomplete team for an agent as follows. Let S be an incomplete team with v_S vacancies. Let the mean value to player i of the unassigned players be μ_i. We then assign the following utility of an incomplete team S to agent i: $\sum_{j\in S} u_i(j) + (v_S - 1)\mu_i$.

Proposition 4. *The One-Player-One-Pick draft is not strategyproof or ex post Pareto efficient.*

[2] The proofs of this and other results are in the supplemental material.

Competitive Equilibrium from Equal Incomes

We now propose a more complex mechanism, based on the *Competitive Equilibrium from Equal Incomes (CEEI)*, which is explicitly designed to achieve allocations that are more ex post fair than the alternative mechanisms.

We begin by defining CEEI, previously introduced by Varian (1974). Given a set of agents N, a set of goods C, and agent preferences over bundles of goods $\succ$, a CEEI mechanism finds a budget $b \in \mathbb{R}_+$ and price vector $p^* \in \mathbb{R}_+^{|C|}$, such that if each agent is allocated its favorite bundle of goods that costs no more than b, then each good in C is allocated to exactly one agent in N, or divided in fractions summing to 1 among the N. In combinatorial allocation problems, such as course allocation, goods (seats in a class) are not divisible, and certain bundles of goods (class schedules) are not allowed to be assigned to an agent. As a result, an exact market clearing tuple (b, p^*) may not exist. To deal with this difficulty, CEEI was relaxed by Budish (2011) to an approximate version, termed A-CEEI. A-CEEI works by assigning nearly equal budgets to all agents, then searching for an approximately market clearing price vector and returning the allocation induced by those prices. The result may not clear the market exactly, but there is an upper bound on the worst-case market clearing error. The resulting allocation satisfies an approximate form of *envy-freeness* (Budish 2011).

Both CEEI and A-CEEI take advantage of the dichotomy between agents and items which agents demand. This makes our setting distinct: agents' demand in team formation is over subsets of other agents. A technical consequence is that this gives rise to a hard constraint for CEEI that if an agent i is paired with agent j, than j must also be paired (assigned to) agent i; any relaxation of this constraint fails to yield a partition on the agents and consequently does not result in an admissible mechanism. We therefore design an approximation of CEEI, termed A-CEEI-TF, that accounts for the specific peculiarities of our setting. Conceptually, the A-CEEI-TF mechanism works by alternating between two steps. First, it searches in price space for approximate relaxed market-clearing prices. Second, it assigns a randomly selected unmatched agent to form a team with its favorite bundle of free agents that is affordable, based on current prices. The result is a mechanism that is strategyproof-in-the-large, and more fair than random serial dictatorship.

A key part of A-CEEI-TF is a *price update function*, which reflects the constraints of the team formation problem. We use a tâtonnement-like price update function f in an auxiliary price space $\tilde{\mathcal{P}} = [-1, 1+\bar{b}]^{|N'|}$, where N' is the set of agents remaining (unassigned) at an iteration of the algorithm, and $\bar{b}$ is the supremum of allowable agent budgets. We make two requirements of a price update function, one ensuring that the iterative updates are well-defined, another to ensure that fixed points of the process are actual solutions.

Definition 1. *A price update function f is* admissible *if (a) its fixed points correspond to (relaxed) market clearing, and (b) $\tilde{\mathcal{P}}$ is closed under f.*

Algorithm 1 A-CEEI-TF Algorithm Outline.

Require: $(N, \succ, \underline{k}, \overline{k})$

1: Randomly assign approximately equal budgets b_i to the agents, $b_i \in [1, \bar{b}]$, $\bar{b} < 1 + 1/|N|$.
2: Randomly order the agents.
3: Search for a price vector p in price space $\mathcal{P} = [0, \bar{b}]^{|N'|}$ that approximately clears the (relaxed) market among the N' remaining agents, given agent budgets b.
4: Take the next unmatched agent in the random order, and assign it to join its favorite bundle of other free agents that it can afford at the current prices, and that leaves a feasible subproblem—i.e., feasible $(N', \underline{k}, \overline{k})$. If the agent cannot afford any remaining bundle of legal size that leaves a feasible subproblem, the agent is assigned its favorite remaining bundle of legal size that leaves a feasible subproblem.
5: Repeat steps 3 and 4 until each agent is on a team.

We now define a candidate price update function, f_{TF}:

$$f_{TF}(\tilde{p})_j = t(\tilde{p})_j + \frac{(1 + \epsilon - (\epsilon/\bar{b})t(\tilde{p})_j)D_j - U_j}{|N'|} \quad (1)$$

where D_j is the number of agents that demand j but whom j does not demand, $U_j = 1$ if and only if no other agent demands j, and 0 otherwise, $\bar{b}$ is the supremum of allowable agent budgets, and $t(\cdot)$ is a truncation function, which takes a price vector $\tilde{p}$ and truncates it to the $[0, \bar{b}]$ interval.

Proposition 5. *$f_{TF}(\cdot)$ is admissible.*

While admissibility of $f_{TF}(\cdot)$ alone does not guarantee convergence of the iterative process, it does guarantee that if convergence happens, we have a solution. The following proposition characterizes some of the properties such solutions possess.

Proposition 6. *A-CEEI-TF is strategyproof-in-the-large. In addition, if A-CEEI-TF yields exact market clearing and induces the same allocation at each stage of price search, it yields envy bounded by a single teammate.*

Proof We sketch a proof of strategyproofness-in-the-large.

If a team formation problem is modified such that each agent is replaced with a measure-one set of copies of itself, each copy being measure zero, we arrive at what is called a continuum economy. If we run A-CEEI-TF in the continuum economy, any individual agent, being zero-measure, has no influence on the approximate equilibrium price vector arrived at by update function $f_{TF}(\cdot)$, at any iteration of the A-CEEI-TF mechanism. Therefore, the only effect the agent can have on the outcome is that, if the agent is randomly selected to choose its favorite affordable team of available agents that leaves a feasible subproblem, the agent's reported preferences determine which team the agent is assigned. Thus, it is a dominant strategy for the agent to report its true preferences, so that in this case the agent will be assigned its most-preferred allowable team. □

Experiments

Although RSD and A-CEEI-TF possess desirable theoretical properties, these results are loose, and the only approximate fairness guarantee, shown for A-CEEI-TF, requires strong assumptions on the environment. We now assess all of the proposed mechanisms empirically through simulations based both on randomly generated classes of preferences, as well as real-world data. Our empirical results turn out to be both one-sided (if one is interested in achieving all three desired properties) and surprising: OPOP, a mechanism with no provable theoretical guarantees, tends to outperform others in fairness, and to perform nearly as well as the best other mechanism in truthfulness and social welfare.

Data Sets: We use both randomly generated data and data from prior studies on preferences of human subjects over each other:

- **Random-similar (R-sim) (Othman, Sandholm, and Budish 2010):** Each agent i, $i \in \{1, 2, \ldots, |N|\}$, is assigned the public value i. A private error term is added to the public value of i to derive the value of i to each other agent j, drawn independently from a normal distribution with zero mean and standard deviation $|N|/5$. The private error term is redrawn until the sum of the private error term and public value is non-negative. Then the value of i to j is the sum of i and the private error term.
- **Random-scattered (R-sca):** In this data set class, the value of a player i is generated independently by each other player j. To determine the value of other players to player j, a total value of 100 is divided at random among the other players as follows. Uniformly random numbers $\in [0, 100]$ are taken, to divide the region into $|N| - 1$ regions. The random draws for agent j are sorted, producing $|N| - 1$ values for the other agents, as the differences between consecutive draws in sorted order.
- **Newfrat:** This data set comes from a widely cited study by Newcomb, in which 17 students at the University of Michigan in 1956 ranked each other in terms of friendship ties. We use the data set from the final week, NEWC15 (Newcomb 1958). We let $\underline{k} = 4$, $\overline{k} = 5$.
- **Freeman:** The data are from a study of email messages sent among 32 researchers in 1978. We use the third matrix of values from the study. The data show how many emails each researcher sent to each other during the study, which we use as a proxy for the strength of directed social links (Freeman and Freeman 1979). We let $\underline{k} = 5$, $\overline{k} = 6$.

For the randomly generated data sets, we set $|N| = 20$ and $\underline{k} = \overline{k} = 5$, and our results are averaged over 20 generated preference rankings for all players.

The four classes of data set we analyze differ most saliently in their number of agents, $|N|$, and in the degree of similarity among the player preferences. For example, Random-similar agents largely agree on which other agents are most valuable, while Random-scattered agents have little agreement. Differences in degree of preference similarity lead to marked differences in the performance outcomes of the various mechanisms.

To measure agent preference similarity in a data set, we let $\mathcal{C}$ equal the mean cosine similarity among all pairs of distinct agents in the data set. Each agent assigns itself a value of 0 or undefined, so we take the cosine similarity between agents i and j only over their values for agents in $N \setminus \{i, j\}$:

$$u_{i-ij} = u_i \setminus \{u_{ii}, u_{ij}\}$$

$$\mathcal{C} = \frac{\sum_{i=1}^{|N|} \sum_{j=i+1}^{|N|} \frac{u_{i-ij} \cdot u_{j-ij}}{\|u_{i-ij}\| \|u_{j-ij}\|}}{(|N|^2 - |N|)/2}$$

In Table 2, we present the mean cosine similarity for each data set we discuss in this paper. Higher cosine similarities indicate greater agreement among agents about the relative values of other agents. We also show the number of agents in each data set.

	$\mathcal{C}$	$\|N\|$	$\underline{k}$	$\overline{k}$
Random-similar 20	0.914	20	5	5
Random-scattered 20	0.499	20	5	5
Newfrat	0.877	17	4	5
Freeman	0.551	32	5	6

Table 2: Mean cosine similarity over all pairs of distinct agents; number of agents; minimum team size; and maximum team size. For random data set classes, $\mathcal{C}$ as shown is the mean over 20 randomly generated instances of the class.

Empirical Analysis of Incentive Compatibility: To study the incentive compatibility of the mechanisms, we used a protocol similar to that used by Vorobeychik and Engel for estimating the regret of a strategy profile (Vorobeychik and Engel 2011). We ran each mechanism on 8 versions of each data set, with different random orders over the players, which we held in common across data sets. We generate deviations from truthful reporting for agent j one at a time until 25 unique deviations have been produced. To produce a deviation from an agent's truthful values for other agents, we first randomly select a number of pairs of values to swap according to a Poisson distribution with $\lambda = 1$, with 1 added. For each pair of values to swap, we first select the rank of one of them, with lower (better) ranks more likely.

The results are shown in Table 3.[3]

RSD is not shown, since it is provably strategyproof, but, remarkably, A-CEEI-TF empirically produces higher (worse) regret of truthful reporting than HBS or OPOP, even though A-CEEI-TF is strategyproof-in-the-large, and the others are not. Both HBS and OPOP appear to offer players only small incentives to lie, with HBS slightly better.

Social Welfare: To facilitate comparison, we normalize the total utility of all teammates for each agent to 1, so that social welfare (already normalized for the number of players) falls in the $[0, 1]$ interval. For comparison, we also include

[3]We only report results for the two random data sets and Newfrat, as it was not feasible to rigorously analyze regret for the far larger Freeman data set. (However, the Freeman data set is similar to Random-scattered; see supplemental material for details.)

	R-sim.	R-sca.	Newfrat
HBS	0.02 ± 0.02	0.04 ± 0.02	0.03 ± 0.02
OPOP	0.07 ± 0.02	0.10 ± 0.05	0.06 ± 0.02
A-CEEI-TF	0.19 ± 0.04	0.29 ± 0.07	0.19 ± 0.03
Max-welfare	0.20 ± 0.03	0.29 ± 0.07	0.22 ± 0.03

Table 3: Mean maximum observed regret of truthful reporting, with 95% confidence intervals.

optimal social welfare for both Random data sets, as well as Newfrat.[4]

The only related theoretical result is that RSD is ex post Pareto optimal; the other three mechanisms do not even possess this guarantee. This makes our results, shown in Tables 4 and 5, remarkable: on Random-similar and Newfrat data sets (both with preferences relatively similar across players), there is little difference in welfare generated by the different mechanisms, but on Random-scattered and Freeman data sets, OPOP statistically significantly outperforms the others.

	R-sim.	R-sca.
RSD	0.22 ± 0.004	0.25 ± 0.01
A-CEEI-TF	0.22 ± 0.004	0.25 ± 0.01
HBS	0.22 ± 0.004	0.25 ± 0.02
OPOP	0.22 ± 0.003	0.27 ± 0.01
Max-welfare	0.25 ± 0.001	0.35 ± 0.01

Table 4: Mean social welfare for the two Random data sets, with 95% confidence intervals.

	Newfrat	Freeman
RSD	0.23 ± 0.01	0.20 ± 0.01
A-CEEI-TF	0.23 ± 0.01	0.19 ± 0.01
HBS	0.22 ± 0.05	0.20 ± 0.01
OPOP	0.22 ± 0.05	0.24 ± 0.02
Max-welfare	0.27 ± 0.00	-

Table 5: Mean social welfare for the Newfrat and Freeman data sets, with 95% confidence intervals.

Fairness: Fairness of an allocation (in our case, a partition of players) can be conceptually described as the relative utility of best- and worst-off agents. Formally, we measure fairness in the experiments as the fraction of agents whose envy is bounded by a single teammate (as defined above).

Our fairness results, shown in Tables 6 and 7 are unambiguous: RSD is always worse, typically by a significant margin, then the other mechanisms. This is intuitive, and is precisely the reason why alternatives to RSD are commonly considered. What is far more surprising is that A-CEEI-TF, in spite of some theoretical promise on the fairness front, and in spite of being explicitly designed for fairness, is in all but one case the *second worst*. While HBS and OPOP are comparable on the high-similarity data sets (Random-similar and Newfrat), it *dominates* all others on the dissimilar data sets (Random-scattered and Freeman).

[4]It was infeasible to compute this for the Freeman data set due to its size.

	R-sim.	R-sca.
RSD	0.43 ± 0.03	0.59 ± 0.05
A-CEEI-TF	0.66 ± 0.05	0.62 ± 0.05
HBS	0.71 ± 0.04	0.61 ± 0.06
OPOP	0.70 ± 0.06	0.79 ± 0.04

Table 6: Mean fraction of agents with envy bounded by a single teammate for the two Random data sets, with 95% confidence intervals.

	Newfrat	Freeman
RSD	0.36 ± 0.02	0.43 ± 0.05
A-CEEI-TF	0.57 ± 0.03	0.55 ± 0.05
HBS	0.67 ± 0.05	0.64 ± 0.04
OPOP	0.68 ± 0.07	0.78 ± 0.05

Table 7: Mean fraction of agents with envy bounded by a single teammate for the Newfrat and Freeman data sets, with 95% confidence intervals.

Next, we consider informal perspectives on the fairness of the various mechanisms. For mechanisms that use a random serial order over players, we can study typical outcomes for a player given its serial index. If players with lower (better) serial indexes receive drastically better outcomes than agents with higher (worse) indexes, such a mechanism is not very fair. In Figure 1 (left), we plot a smoothed version of the mean fraction of total utility achieved by agents at each random serial index from 0 to 19, for the mechanisms RSD, HBS draft, OPOP draft, and A-CEEI-TF. Results are based on 20 instances of Random-scattered preferences, held in common across the mechanisms, with different serial orders over the players. From this Figure, it is apparent that random serial dictatorship gives far better outcomes to the best-ranked agents than to any others. Surprisingly, A-CEEI-TF follows a similar pattern in this case, although we did observe that for some other game types (not shown), A-CEEI-TF's curve gives better outcomes to low-ranked agents than RSD. In the HBS draft, a "shelf" of high utility for the several best-ranked agents is typical, as all of the team captains receive similarly high utility, with a steep drop-off in utility for non-captain agents. In the OPOP draft, in contrast to all others, the utility curve is far more flat across random serial indexes: even the agents with high (bad) serial indexes achieve moderately good outcomes for themselves.

Some mechanisms for team formation tend to give better outcomes to an agent that is "popular," having a high mean value to the other agents. For example, random serial dictatorship biases outcomes in favor of popular players, because even if a popular player is not a team captain, it is likely that this player will be selected by some team captain along with other desirable players. An unpopular player, however, will likely be left until near the final iteration of RSD, to be selected along with other unpopular players. Therefore,

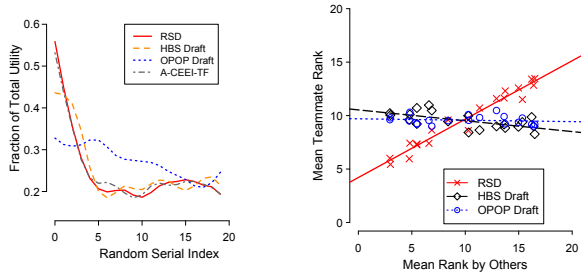

Figure 1: Left: Mean fraction of total utility earned versus the random serial index of the agent. Fraction of total utility is between 0 (worst) and 1 (best). A cubic smoothing spline is applied. Right: Mean rank of an agent's teammates, versus mean rank of the agent by other agents. Possible ranks range from 1 (best) to 20 (worst). Each point represents a single agent's mean outcome. Best-fit lines use ordinary least squares.

we might expect RSD to yield better outcomes to popular players, especially when agents' preferences are highly similar. To quantify this intuition, we plot in Figure 1 (right) the mean rank of an agent's teammates according to the agent's preferences, versus the agent's mean rank assigned by the other agents. Each point in the scatter plot represents a single agent's mean outcomes across 20 instances of Random-similar preferences, held in common across the mechanisms. We find best-fit lines via OLS regression, for each of RSD, HBS draft, and OPOP draft. The results indicate that, as expected, RSD offers better outcomes to popular agents than to unpopular ones, with a distinctly positive trend line. The HBS draft and OPOP draft appear less biased for or against popular agents, with OPOP showing slightly lower correlation than HBS between an agent's popularity and the mean value of its assigned team.

Conclusion

We considered team formation as a mechanism design problem, in which the mechanism elicits agents' preferences over potential teammates in order to partition the agents into teams. The teams produced should have high social welfare and fairness, in the sense that few agents should prefer to switch teams with others. We proposed two novel mechanisms for this problem: a version of approximate competitive equilibrium for equal incomes (A-CEEI-TF), and the one-player-one-pick draft (OPOP). We showed theoretically that A-CEEI-TF is strategyproof-in-the-large and approximates envy-freeness. OPOP lacks these theoretical guarantees but empirically outperformed A-CEEI-TF in truthfulness and fairness, as well as in social welfare for data sets with sufficiently dissimilar agent preferences. In addition, OPOP surpassed other mechanisms tested, including random serial dictatorship and the HBS draft, in social welfare and fairness. The HBS draft, however, produced slightly better truthfulness that the OPOP draft. Given the relative simplicity of implementing OPOP, this mechanism emerges as a strong candidate for team formation settings.

References

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

Alcalde, J., and Revilla, P. 2004. Researching with whom? Stability and manipulation. *Journal of Mathematical Economics* 40(8):869–887.

Aumann, R. J., and Dreze, J. H. 1974. Cooperative games with coalition structures. *International Journal of Game Theory* 3(4):217–237.

Aziz, H.; Brandt, F.; and Seedig, H. G. 2011. Stable partitions in additively separable hedonic games. In *The 10th International Conference on Autonomous Agents and Multiagent Systems*, 183–190.

Ballester, C. 2004. NP-completeness in hedonic games. *Games and Economic Behavior* 49(1):1–30.

Banerjee, S.; Konishi, H.; and Sönmez, T. 2001. Core in a simple coalition formation game. *Social Choice and Welfare* 18(1):135–153.

Bogomolnaia, A., and Jackson, M. O. 2002. The stability of hedonic coalition structures. *Games and Economic Behavior* 38(2):201–230.

Budish, E. B., and Cantillon, E. 2012. The multi-unit assignment problem: Theory and evidence from course allocation at Harvard. *American Economic Review* 102(5):2237–2271.

Budish, E. 2011. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy* 119(6):1061–1103.

Budish, E. 2012. Matching "versus" mechanism design. *ACM SIGECOM Exchanges* 11(2):4–15.

Cechlarova, K., and Romero-Medina, A. 2001. Stability in coalition formation games. *International Journal of Game Theory* 29(4):487–494.

Freeman, S. C., and Freeman, L. C. 1979. FreemansEIES: Weighted static one-mode network (messages). http://toreopsahl.com/datasets/#FreemansEIES.

Li, C.; Chawla, S.; Rajan, U.; and Sycara, K. 2004. Mechanism design for coalition formation and cost sharing in group-buying markets. *Electronic Commerce Research and Applications* 3:341–354.

Marcolino, L. S.; Jiang, A. X.; and Tambe, M. 2013. Multi-agent team formation: Diversity beats strength? In *International Joint Conference on Artificial Intelligence*, 279–285.

Newcomb, N. 1958. Newcomb, Nordlie: Fraternity. http://moreno.ss.uci.edu/data.html.

Othman, A.; Sandholm, T.; and Budish, E. 2010. Finding approximate competitive equilibria: Efficient and fair course allocation. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, 873–880.

Roth, A. E., and Peranson, E. 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review* 89(4):748–780.

Sung, S.-C., and Dimitrov, D. 2010. Computational complexity in additive hedonic games. *European Journal of Operational Research* 203(3):635–639.

Varian, H. 1974. Equity, envy and efficiency. *Journal of Economic Theory* 29(2):217–244.

Vorobeychik, Y., and Engel, Y. 2011. Average-case analysis of VCG with approximate resource allocation algorithms. *Decision Support Systems* 51(3):648–656.