Games and Economic Behavior 82 (2013) 562–581

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Pareto optimality in coalition formation

Haris Aziz [a], Felix Brandt [b,*], Paul Harrenstein [c]

[a] NICTA and University of New South Wales, 2033 Sydney, Australia
[b] Technische Universität München, 80538 München, Germany
[c] University of Oxford, Oxford OX1 3QD, United Kingdom

ARTICLE INFO

Article history:
Received 8 December 2011
Available online 3 September 2013

JEL classification:
C63
C70
C71
C78

Keywords:
Coalition formation
Hedonic games
Pareto optimality
Computational complexity

ABSTRACT

A minimal requirement on allocative efficiency in the social sciences is Pareto optimality. In this paper, we identify a close structural connection between Pareto optimality and perfection that has various algorithmic consequences for coalition formation. Based on this insight, we formulate the *Preference Refinement Algorithm* (*PRA*) which computes an individually rational and Pareto optimal outcome in hedonic coalition formation games. Our approach also leads to various results for specific classes of hedonic games. In particular, we show that computing and verifying Pareto optimal partitions in general hedonic games, anonymous games, three-cyclic games, room-roommate games and B-hedonic games is intractable while both problems are tractable for roommate games, W-hedonic games, and house allocation with existing tenants.

1. Introduction

Ever since the publication of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* in 1944, coalitions have played a central role within game theory. The crucial questions in coalitional game theory are which coalitions can be expected to form and how the members of coalitions should divide the proceeds of their cooperation. Traditionally the focus has been on the latter issue, which led to the formulation and analysis of concepts such as the core, the Shapley value, or the bargaining set. Which coalitions are likely to form is commonly assumed to be settled exogenously, either by explicitly specifying the coalition structure, a partition of the players in disjoint coalitions, or, implicitly, by assuming that larger coalitions can invariably guarantee better outcomes to its members than smaller ones and that, as a consequence, the grand coalition of all players will eventually form. The two questions, however, are clearly interdependent: the individual players' payoffs depend on the coalitions that form just as much as the formation of coalitions depends on how the payoffs are distributed.

Coalition formation games, which were first formalized by Drèze and Greenberg (1980), model coalition formation in settings in which utility is *non-transferable*. In many such situations it is natural to assume that a player's appreciation of a coalition structure only depends on the coalition he is a member of and not on how the remaining players are grouped. Initiated by Banerjee et al. (2001) and Bogomolnaia and Jackson (2002), much of the work on coalition formation now concentrates on these so-called *hedonic games*. In this paper, we focus on Pareto optimality and individual rationality in this rich class of coalition formation games.

* Corresponding author.
E-mail addresses: haris.aziz@nicta.com.au (H. Aziz), brandtf@in.tum.de (F. Brandt), paul.harrenstein@cs.ox.ac.uk (P. Harrenstein).

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2013.08.006

The main question in coalition formation games is which coalitions one may reasonably expect to form. To get a proper formal grasp of this issue, a number of stability concepts have been proposed for hedonic games—such as the core or Nash stability—and much research concentrates on conditions for existence, the structure, and computation of stable and efficient partitions. *Pareto optimality*—which holds if no coalition structure is strictly better for some player without being strictly worse for another—and *individual rationality*—which holds if no player would rather be on his own—are commonly considered minimal requirements for any reasonable partition.[1]

Another reason to investigate Pareto optimal partitions algorithmically is that, in contrast to other stability concepts like the core, they are guaranteed to exist. This even holds if we additionally require individual rationality. Moreover, while core stable matchings in marriage games are guaranteed to exist and can be found efficiently using the *Gale–Shapley algorithm*, checking their existence in almost any noteworthy generalization such as roommate games (Ronn, 1990), general hedonic games (Ballester, 2004), and games with $\mathscr{B}$- and $\mathscr{W}$-preferences (Cechlárová and Hajduková, 2004a, 2004b) is NP-hard. Interestingly, when the status-quo partition cannot be changed without the mutual consent of all players, Pareto optimality can be seen as a notion of stability (Morrill, 2010).

When there are indifferences in the preferences, a core stable outcome is not necessarily Pareto optimal. Thus, Pareto optimality can serve as a refinement of core stable outcomes. An outcome which is Pareto optimal and a Pareto improvement over a core stable outcome is called *Pareto-stable*. This notion further motivates the need for algorithms to compute Pareto improvements of given outcomes. Sotomayor and Özak (2009) note that "*the study of the discrete two-sided matching models with non-necessarily strict preferences and the search for algorithms to produce the Pareto-stable matchings is a new and interesting line of investigation*."

We investigate both the problem of finding a Pareto optimal and individually rational partition and the problem of deciding whether a given partition is Pareto optimal. In particular, our results concern *general hedonic games*, *B-hedonic* and *W-hedonic games* (two classes of games in which each player's preferences over coalitions are based on his most preferred and least preferred player in his coalition, respectively), *roommate games*, *house allocation with existing tenants*, *three-cyclic games*, *room-roommate games*, and *anonymous games*.

Many of our results, both positive and negative, rely on the concept of *perfection* and how it relates to Pareto optimality. A *perfect* partition is one that is most desirable for every player. We find (a) that under extremely mild conditions, NP-hardness of finding a perfect partition implies NP-hardness of finding a Pareto optimal partition (Lemma 1), and (b) that under stronger but equally well-specified circumstances, feasibility of finding a perfect partition implies feasibility of finding a Pareto optimal partition (Lemma 2). The latter we show via a Turing reduction to the problem of computing a perfect partition. At the heart of this algorithm, which we refer to as the *Preference Refinement Algorithm* (*PRA*), lies a fundamental insight of how perfection and Pareto optimality are related. It turns out that a partition is Pareto optimal for a particular preference profile if and only if the partition is perfect for another but related profile (Theorem 1). In this way PRA is also applicable to any other discrete allocation setting.

A well-established procedure for finding Pareto optimal allocations is *serial dictatorship* in which agents are invoked according to some fixed order and each agent subsequently narrows down the set of possible allocations to his most preferred ones (see, e.g., Abdulkadiroğlu and Sönmez, 1998). When applied to compactly represented coalition formation games, serial dictatorship can be extremely inefficient from a computational point of view even when preferences over coalitions are strict. Moreover, there can be Pareto optimal partitions that serial dictatorship is unable to find, which may have serious repercussions if also other considerations, like fairness, are taken into account. By contrast, PRA handles indifferences better and is complete in the sense that it may return any Pareto optimal partition, provided that the subroutine that computes perfect partitions can compute any perfect partition (Theorem 3). PRA has also been designed to compute Pareto optimal Pareto improvements over a given outcome which is an important problem in resource allocation and coalition formation.

2. Preliminaries

In this section, we review the terminology and notation used in this paper.

Hedonic games. Let N be a set of n players. A *coalition* is a non-empty subset of N. By $\mathscr{N}_i$ we denote the set of coalitions player i belongs to, i.e., $\mathscr{N}_i = \{S \subseteq N\colon\ i \in S\}$. A *coalition structure*, or simply a *partition*, is a partition π of the players N into coalitions, where $\pi(i)$ is the coalition player i belongs to.

A *hedonic game* is a pair (N, R), where $R = (R_1, \dots, R_n)$ is a *preference profile* specifying the preferences of each player i as a binary, complete, reflexive, and transitive *preference relation* R_i over $\mathscr{N}_i$. If R_i is also anti-symmetric we say that i's preferences are *strict*. We adopt the conventions of social choice theory by writing $S\ P_i\ T$ if $S\ R_i\ T$ but not $T\ R_i\ S$—i.e., if i *strictly prefers* S to T—and $S I_i T$ if both $S\ R_i\ T$ and $T\ R_i\ S$—i.e., if i is *indifferent* between S and T.

For a player i, a coalition S in $\mathscr{N}_i$ is *acceptable* if for i being in S is at least preferable as being alone—i.e., if $S\ R_i\ \{i\}$—and *unacceptable* otherwise.

[1] For example, in the context of TU coalitional games, Aumann (1987) states that "the requirement that a feasible outcome y be undominated via one-person coalitions (individual rationality) and via the all-person coalition (efficiency or Pareto optimality) is thus quite compelling." His point can easily be seen to extend to hedonic games as well.