Proceedings, Fifteenth International Conference on
Principles of Knowledge Representation and Reasoning (KR 2016)

Boolean Hedonic Games

Haris Aziz
Data61 and University
of New South Wales
Australia

Paul Harrenstein
Dept. of Computer Science
University of Oxford
United Kingdom

Jérôme Lang
LAMSADE
Université Paris-Dauphine
France

Michael Wooldridge
Dept. of Computer Science
University of Oxford
United Kingdom

Abstract

We study *hedonic games with dichotomous preferences*. Hedonic games are cooperative games in which players desire to form coalitions, but only care about the makeup of the coalitions of which they are members; they are indifferent about the makeup of other coalitions. The assumption of dichotomous preferences means that, additionally, each player's preference relation partitions the set of coalitions of which that player is a member into just two equivalence classes: satisfactory and unsatisfactory. A player is indifferent between satisfactory coalitions, and is indifferent between unsatisfactory coalitions, but strictly prefers any satisfactory coalition over any unsatisfactory coalition. We develop a succinct representation for such games, in which each player's preference relation is represented by a propositional formula. We show how solution concepts for hedonic games with dichotomous preferences are characterised by propositional formulas.

Introduction

Hedonic games are cooperative games in which players desire to form coalitions, but only care about the makeup of the coalitions of which they are members; they are indifferent about the makeup of other coalitions. Drèze and Greenberg (1980) suggested that coalition formation involves a *hedonic aspect*, i.e., that apart from the yield of a coalition, players may also be interested in its composition. Bogomolnaia and Jackson (2002) and Banerjee, Konishi, and Sönmez (2001) then defined hedonic games in their present form as a simple but very versatile model of coalition formation. Hedonic games capture many social, political, and economic group formation scenarios, and can be seen as a generalisation of the stable marriage setting (Aziz and Savani 2016).

As the specification of a hedonic game requires the expression of each player's ranking over all sets of players including him, in general, such a specification requires exponential space – and, when used by a centralised mechanism, exponential elicitation time. Such an exponential blow-up severely limits the practical applicability of hedonic games, and for this reason researchers have investigated compactly represented hedonic games. One approach to this problem has been to consider possible restrictions on the possible preferences that players have. For example, one may assume that each player specifies only a ranking over single players, and that her preferences over coalitions are defined according to the identity of the best (respectively, worst) element of the coalition (Cechlárová and Hajduková 2004; Cechlárová 2008). One may also assume that each player's preferences depend only on the number of players in her coalition (Bogomolnaia and Jackson 2002). These representations come with a domain restriction, i.e., a loss of expressivity. Elkind and Wooldridge (2009) consider a fully expressive representation for hedonic games, based on weighted logical formulas; in the worst case, their representation requires space exponential in the number of players, but in many cases the space requirement is much smaller.

In this paper, we consider another natural restriction on player preferences. We consider hedonic games with *dichotomous preferences*. The assumption of dichotomous preferences means that each player's preference relation partitions the set of coalitions of which that player is a member into just two equivalence classes: satisfactory and unsatisfactory. A player is indifferent between satisfactory coalitions, and is indifferent between unsatisfactory coalitions, but strictly prefers any satisfactory coalition over any unsatisfactory coalition.

Dichotomous preferences have been studied in other economic settings, such as by Bogomolnaia, Moulin, and Stong (2005), Bogomolnaia and Moulin (2004), and Bouveret and Lang (2008) in the context of fair division, by Harrenstein et al. (2001) in the context of Boolean games, by Konieczny and Pino-Pérez (2002) in the context of belief merging, by Bogomolnaia and Moulin (2004) in the context of matching, and by Brams and Fishburn (2007) (and many others) in the context of approval voting.

When the space of all possible alternatives has a combinatorial structure, propositional formulas are a very natural representation of dichotomous preferences. In such a representation, variables correspond to goods (in fair division), outcome variables (Boolean games), state variables (belief merging), or players (coalition formation). In the latter case, which we will be concerned with in the present paper, each player i can express her preferences over coalitions containing her by using propositional atoms of the form ij ($j \neq i$), meaning that i and j are in the same coalition. Thus, for example, player 1 can express by the formula $(12 \vee 13) \wedge \neg 14$ that he wants to be in a coalition with player 2 or with player 3, but not with player 4. Our primary aim is to present such a propositional framework for specifying hedonic games and computing various solution concepts. We will first define a propositional logic using atoms of the form ij, together with domain axioms expressing that the output of the game should be a partition of the set of players. Then we consider a range of solution concepts, and show that they can be characterised by specific classes of formulas, and solved using propositional satisfiability solvers. The result is a simple, natural, and compact representation scheme for expressing preferences, and, as our characterisations are model preserving, a machinery based on satisfiability for computing partitions satisfying some specific stability criteria such as Nash stability or core stability.

Preliminaries

In this section, we recall some definitions relating to coalitions, coalition structures (or partitions), and hedonic games. See, e.g., (Chalkiadakis, Elkind, and Wooldridge 2011) for an in-depth discussion of these and related concepts.

Coalitions and Partitions We consider a setting in which there is a set N of n players with typical elements $i, j, k, \dots$. Players can form *coalitions*, which we will denote by $S, T, \dots$. A coalition is a subset of the players N. One may usefully think of the players as getting together to form teams that will work together. A *coalition structure* is an exhaustive partition $\pi = \{S_1, \dots, S_m\}$ of the players into disjoint coalitions, i.e., $S_1 \cup \cdots \cup S_m = N$ and $S_i \cap S_j = \emptyset$ for all $S_i, S_j \in \pi$ such that $i \neq j$. For technical convenience, we slightly deviate from standard conventions and require that every coalition structure π contains the empty set $\emptyset$. We commonly refer to coalition structures simply as *partitions*. In examples, we also write, e.g., $[\,12\,|\,34\,|\,5\,]$ rather than the more cumbersome $\{\{1,2\},\{3,4\},\{5\},\emptyset\}$, and $\{234\}$ rather than $\{2,3,4\}$. For each player i in N, we let $\mathscr{N}_i = \{S \subseteq N : i \in S\}$ denote the set of coalitions over N that contain i. If π is a partition, then $\pi(i)$ refers to the coalition in π that player i is a member of.

The notion of players leaving their own coalition and joining another lies at the basis of many of the solution concepts that we will come to consider. We introduce some notation to represent such situations. For T a group of players (not necessarily a coalition in π), by $\pi|_T$ we refer to the partition $\{S_1 \cap T, \dots, S_m \cap T\}$ and we write $\pi|_{-T}$ for $\pi|_{N \setminus T}$. Moreover, for S a coalition in partition $\pi|_{-T}$, we use $\pi[T \to S]$ to refer to the partition that results if the players in T leave their respective coalitions in π and join coalition S. We also allow T to form a coalition of its own, in which case we write $\pi[T \to \emptyset]$. Formally, we have, for $S \in \pi|_{-T}$,

$$\pi[T \to S] = \{S_j \in \pi|_{-T} : S_j \neq S\} \cup \{S \cup T, \emptyset\}.$$

If T is a singleton $\{i\}$ we also write $\pi|_{-i}$ and $\pi[i \to S]$ instead of $\pi|_{-\{i\}}$ and $\pi[\{i\} \to S]$, respectively. Thus, e.g., $S \cup \{i\} \in \pi[i \to S]$ and $\pi[i \to \pi(i) \setminus \{i\}] = \pi$.

Finally, we define $\pi[i \rightleftarrows j]$ as the partition where i and j exchange their places, i.e., $\pi[i \rightleftarrows j]$ is given by:

$$(\pi \setminus \{\pi(i), \pi(j)\}) \cup \{(\pi(i) \setminus \{i\}) \cup \{j\}, (\pi(j) \setminus \{j\}) \cup \{i\}\}.$$

Thus, for partition $\pi = [123|45]$, we have $\pi(1) = \pi(2) = \{123\}$ and $\pi(4) = \{45\}$. Furthermore, $\pi|_{\{1245\}} = [12|45]$ and $\pi|_{-\{34\}} = [12|5]$. Also, $\pi[1 \to \{45\}] = [23|145]$, $\pi[1 \to \emptyset] = [1|23|45]$, $\pi[\{14\} \to \emptyset] = [23|5|14]$ and $\pi[3 \rightleftarrows 4] = [124|35]$.

Hedonic Games Hedonic games are the class of coalition formation games in which each player is only interested in the coalition he is a member of, and is indifferent as to how the players outside his own coalition are grouped. Hedonic games were originally introduced by Drèze and Greenberg (1980) and further developed by, e.g., Bogomolnaia and Jackson (2002). Also see (Hajduková 2006) and (Aziz and Savani 2016) for a survey from a more computational point of view. Formally, a *hedonic game* is a tuple $(N, \succeq_1, \dots, \succeq_n)$, where $\succeq_i$ represents i's transitive, reflexive, and complete preferences over the set of all coalitions $\mathscr{N}_i$ containing i. Thus, $S \succeq_i T$ intuitively signifies that player i considers coalition S at least as desirable as coalition T, where S and T are coalitions in $\mathscr{N}_i$. By $\succ_i$ we denote the strict part of $\succeq_i$. The preferences $\succeq_i$ of a player i are said to be *dichotomous* whenever $\mathscr{N}_i$ can be partitioned into two disjoint sets $\mathscr{N}_i^+$ and $\mathscr{N}_i^-$ such that i strictly prefers all coalitions in $\mathscr{N}_i^+$ to those in $\mathscr{N}_i^-$ and is indifferent otherwise, i.e., $S \succ_i T$ if and only if $S \in \mathscr{N}_i^+$ and $T \in \mathscr{N}_i^-$. A coalition S in $\mathscr{N}_i$ is *acceptable* to i if i (weakly) prefers S to coalition $\{i\}$, where he is on his own, i.e., if $S \succeq_i \{i\}$. By contrast, we say that a coalition S is *satisfactory* or *desirable* for i if $S \in \mathscr{N}_i^+$. Satisfactory partitions are thus generally acceptable to all players. The implication in the other direction, however, does not hold.

Because every player is indifferent as to how the players outside his own coalition are grouped, preferences on coalitions are lifted to preferences on partitions: player i prefers partition π to partition π' whenever i prefers coalition $\pi(i)$ to coalition $\pi'(i)$. We also extend the concepts of acceptability and desirability of coalitions to partitions.

Example 1 *Consider the following hedonic game with four players,* 1, 2, 3, *and* 4, *whose (dichotomous) preferences are as follows. (Indifferences are indicated by commas.)*

$$1\colon \{123\}, \{124\}, \{134\}, \{1234\} \succ_1 \{1\}, \{12\}, \{13\}, \{14\}$$
$$2\colon \{213\}, \{214\}, \{234\} \succ_2 \{2\}, \{21\}, \{23\}, \{24\}, \{2134\}$$
$$3\colon \{31\}, \{32\}, \{312\} \succ_3 \{3\}, \{34\}, \{314\}, \{324\}, \{3124\}$$
$$4\colon \{41\}, \{42\}, \{43\}, \{412\}, \{413\}, \{4\} \succ_4 \{423\}, \{4123\}$$

Thus, player 1 *(resp.* 2*) wants to be in a coalition of at least (resp. exactly) three. Player* 3 *wants to be in the same coalition as* 1 *or as* 2*, but not together with* 4*. Player* 4 *does not want to be with players* 2 *and* 3 *together. There is exactly one partition that is satisfactory for all four players, namely* $[123\,|\,4]$*. For players* 1, 2, 3*, all coalitions are acceptable. For player* 4*,* $\{423\}$ *and* $\{1234\}$ *are unacceptable.*

Solution Concepts for Hedonic Games A *solution concept* associates with every hedonic game $(N, \succeq_1, \dots, \succeq_n)$ a (possibly empty) set of partitions of N. Here we review some of the most common solution concepts.

Individual rationality captures the idea that every player prefers the coalition he is in to being on his own, i.e., that coalitions are acceptable to its members. Thus, formally, π is *individually rational* if, for all players i in N,

$$\pi(i) \succeq_i \{i\}.$$

This condition is obviously equivalent to $\pi \succeq_i \pi[i \to \emptyset]$.

For dichotomous hedonic games, a partition π is said to be *social welfare optimal* if it maximises the number of players who are in a satisfactory coalition, i.e., if π maximises $|\{i \in N : \pi(i) \in \mathscr{N}_i^+\}|$. In a similar way, a partition π is *Pareto optimal* if it maximises the set of players being in a satisfactory coalition with respect to set-inclusion, i.e., if there is no partition π' with

$$\{i \in N : \pi(i) \in \mathscr{N}_i^+\} \subsetneq \{i \in N : \pi'(i) \in \mathscr{N}_i^+\}.$$

In the extreme case in which every player is in a most preferred coalition, π is said to be *perfect* (Aziz, Brandt, and Harrenstein 2013). A perfect partition satisfies any other of our stability concepts.

A partition is *Nash stable* if no player would like to unilaterally abandon the coalition he is in and join any other existing coalition or stay on his own, i.e., if, for all $i \in N$ and all $S \in \pi$,

$$\pi(i) \succeq_i S \cup \{i\}.$$

Observe that this condition is equivalent to $\pi \succeq_i \pi[i \to S]$.

Core stability concepts consider group deviations instead of individual ones. A group of players, possibly from different coalitions, is said to block a partition if they would all benefit by joining together in a separate coalition. Formally, T *blocks* (or *is blocking*) partition π if, for all $i \in T$,

$$T \succ_i \pi(i).$$

Thus, T blocks π if and only if $\pi[T \to \emptyset] \succ_i \pi$ for all $i \in T$. A group T *weakly blocks* (or *is weakly blocking*) π if $T \succeq_i \pi(i)$ holds for all $i \in T$ and $T \succ_i \pi(i)$ holds for some $i \in T$. Then, π is *core stable* if no group is blocking it and π is *strict core stable* if no group is weakly blocking it.

Partition π is envy-free if no player is envious of another player, i.e., if no player i would prefer to change places with another player j. Formally, partition π is *envy-free* if, for all players i and j,

$$\pi \succeq_i \pi[i \leftrightarrows j].$$

If $\pi[i \leftrightarrows j] \succ_i \pi$ we also say that player i *envies* player j.

Example 1 (continued) *In our example, in partition* $[123 \,|\, 4]$ *each player is in a most preferred coalition. As such* $[123 \,|\, 4]$ *is perfect as well as social welfare optimal and satisfies all solution concepts mentioned above. All partitions except* $[1 \,|\, 234]$ *and* $[1234]$ *are individually rational.*

Now, consider partition $\pi = [1 \,|\, 23 \,|\, 4]$*. Here, player* 2 *does not want to abandon her coalition* $\{23\}$ *and join another as she prefers none of the following partitions to* π*:* $\pi[2 \to \{1\}] = [12 \,|\, 3 \,|\, 4]$*,* $\pi[2 \to \{23\}] = [1 \,|\, 23 \,|\, 4]$*,* $\pi[2 \to \{4\}]$*, and* $\pi[2 \to \emptyset] = [1 \,|\, 2 \,|\, 3 \,|\, 4]$*. As, however,* $\pi[1 \to \{23\}] = [123 \,|\, 4]$ *and* $[123 \,|\, 4] \succ_1 \pi$*, partition* π *is not Nash stable.*

Also observe that for $\pi = [1 \,|\, 23 \,|\, 4]$ *the group* $\{124\}$ *is strongly blocking, as* $\pi[\{124\} \to \emptyset] = [124 \,|\, 3]$ *and* $[124 \,|\, 3] \; P_i \; \pi$ *for all* $i \in \{124\}$*. Thus,* π *is not core stable. By contrast,* $[14 \,|\, 23]$ *is core stable: players* 3 *and* 4 *are satisfied and thus do not want to deviate, while players* 1 *and* 2 *cannot form a blocking coalition without* 3 *or* 4*. However,* $\{124\}$ *is still weakly blocking, and as such* $[14 \,|\, 23]$ *is not strict core stable.*

For envy-freeness, consider partition $\pi' = [1 \,|\, 24 \,|\, 3]$*. Then, player* 3 *envies player* 4*, as* $\pi'[3 \leftrightarrows 4] = [1 \,|\, 23 \,|\, 4]$ *and* $[1 \,|\, 23 \,|\, 4] \succ_3 \pi'$*. By contrast, player* 3 *does not envy player* 2*: we have* $\pi'[3 \leftrightarrows 2] = [1 \,|\, 2 \,|\, 34]$ *but not* $[1 \,|\, 2 \,|\, 34] \succ_3 \pi'$*.*

A Logic for Coalition Structures

In this section, we develop a logic for representing coalition structures. We will then use this logic as a compact specification language for dichotomous preference relations in hedonic games.

Syntax Given a set N of n players, we define a propositional language L_N built from the classical connectives and containing for every (unordered) pair $\{i, j\}$ of distinct players a propositional variable $p_{\{i,j\}}$. The set of propositional variables we denote by V. Observe that $|V| = \binom{n}{2}$. For notational convenience we will write ij for $p_{\{i,j\}}$. Thus, ij and ji refer to the same symbol. The language is interpreted on coalition structures on N and the informal meaning of ij is "i and j are in the same coalition." Formally, the formulas of the language L_N, with typical element φ is given by the following grammar

$$\varphi ::= ij \mid \neg\varphi \mid (\varphi \vee \varphi)$$

where $i, j \in N$ and $i \neq j$. By $|\varphi|$ we denote the *size* of φ. The classical connectives $\bot$, $\top$, $\wedge$, $\to$, and $\leftrightarrow$ are defined in the usual way. For i a player, we write V_i for the propositional variables in which i appears, i.e.,

$$V_i = \{ij \in V : j \in N \setminus \{i\}\}.$$

Note that for distinct players i and j we have $V_i \cap V_j = \{ij\}$. With a slight abuse of notation, we denote the propositional language over V_i by L_i. We also make use of the following useful notational shorthand:

$$i_1 \cdots i_m \bar{i}_{m+1} \cdots \bar{i}_p = \bigwedge_{1 \leq j \leq m} i_1 i_j \wedge \bigwedge_{m < k \leq p} \neg i_1 i_k.$$

Thus, $i_1 \cdots i_m \bar{i}_{m+1} \cdots \bar{i}_p$ conveys that $i_1, \ldots, i_m$ are in the same coalition and each of them in another coalition than $i_{m+1} \cdots i_p$. Thus, where $N = \{1, 2, 3, 4\}$, $12\overline{34} \vee 13\overline{24} \vee 14\overline{23}$ abbreviates $(12 \wedge \neg 13 \wedge \neg 14) \vee (13 \wedge \neg 12 \wedge \neg 14) \vee (14 \wedge \neg 12 \wedge \neg 13)$ and signifies that player 1 is in a coalition of two players.

Peters (2016a; 2016b) uses a slightly different language than we, where in the goal expressed by player i, i is left implicit. For instance, if the goal of agent 1 is expressed in our language by $12 \wedge \neg 13$ (and abbreviated into $12\bar{3}$) then it would be expressed in his language by $2 \wedge \neg 3$. Obviously, both languages are almost identical, and the choice of either of them has no impact on any of the results.

Semantics We interpret the formulas of L_N on partitions π as follows.

$$\begin{array}{lll} \pi \models ij & \text{iff} & \pi(i) = \pi(j) \\ \pi \models \neg\varphi & \text{iff} & \pi \not\models \varphi \\ \pi \models \varphi \vee \psi & \text{iff} & \pi \models \varphi \text{ or } \pi \models \psi \end{array}$$

For $\Psi \subseteq L_N$, we have $\Psi \models \varphi$ if $\pi \models \psi$ for all $\psi \in \Psi$ implies $\pi \models \varphi$. If $\Psi = \emptyset$, we write $\models \varphi$ and say that φ is *valid*.

Notice that partitions play a dual role in our framework: both their initial role as coalition structures, and the role of models in our logic. This dual role is key to using formulas of our propositional language as a specification language for preference relations. Thus, e.g., partition $[1|2|345]$ satisfies the following formulas of L_N: 345, $3\overline{1}$, $345\overline{12}$, $\neg 12 \wedge (23 \vee 34)$, and $12 \leftrightarrow 23$.

Axiomatisation We have the following axiom schemes for mutually distinct players i, j, and k,

(A0) all propositional tautologies

(A1) $ij \wedge jk \rightarrow ik$ (*transitivity*)

as well as *modus ponens* as the only rule of the system:

(MP) from φ and $\varphi \rightarrow \psi$ infer ψ. (*modus ponens*)

The resulting logic we refer to as $\mathbf{P}$ and write $\Psi \vdash_{\mathbf{P}} \varphi$ if there is a derivation of φ from Ψ, (A0), and (A1), using modus ponens.

Proposition 1 (Completeness) *Let $\Psi \cup \{\varphi\} \subseteq L_N$. Then,*

$$\Psi \vdash_{\mathbf{P}} \varphi \quad \textit{iff} \quad \Psi \models \varphi.$$

Proof sketch: Soundness is straightforward. For completeness a standard Lindenbaum construction can be used. To this end, assume $\Psi \nvdash_{\mathbf{P}} \varphi$. Then, $\Psi \cup \{\neg\varphi\}$ is consistent and can as such be extended to a maximal consistent theory Ψ^*. Define a relation $\sim_{\Psi^*}$ such that for all $i, j \in N$,

$$i \sim_{\Psi^*} j \quad \text{iff} \quad ij \in \Psi^*.$$

Together with (MP), the axiom schemes (A0) and (A1) ensure that $\sim_{\Psi^*}$ is a well-defined equivalence relation. Let $[i]_{\sim_{\Psi^*}} = \{j \in N : i \sim_{\Psi^*} j\}$ be the equivalence class under $\sim_{\Psi^*}$ to which player i belongs. Then define the partition $\pi_{\Psi^*} = \{[i]_{\sim_{\Psi^*}} : i \in N\}$. By a straightforward structural induction, it can then be shown that for all $\psi \in L_N$,

$$\pi_{\Psi^*} \models \psi \quad \text{iff} \quad \psi \in \Psi^*.$$

It follows that $\pi_{\Psi^*} \models \Psi$ and $\pi_{\Psi^*} \not\models \varphi$. Hence, $\Psi \not\models \varphi$. □

As an aside, note that one can reason with coalition structures in standard propositional logic, by writing the transitivity axiom directly as a propositional logic formula. Let

$$\mathit{trans} = \bigwedge_{i,j,k \in N} (ij \wedge jk \rightarrow ik),$$

where i, j, and k are assumed to be distinct. Then, for all propositional formulas φ and ψ of L_N,

$$\varphi \vdash_{\mathbf{P}} \psi \quad \text{iff} \quad \varphi \wedge \mathit{trans} \vdash \psi,$$

i.e., checking whether a formula φ implies another formula ψ in $\mathbf{P}$ is equivalent to saying that φ together with the transitivity constraint implies ψ. This means that reasoning tasks in $\mathbf{P}$ can be done with a classical propositional theorem prover. In what follows we say that two formulas φ and ψ are $\mathbf{P}$-equivalent whenever their equivalence can be proven in $\mathbf{P}$, i.e., $\vdash_{\mathbf{P}} \varphi \leftrightarrow \psi$.

Boolean Hedonic Games

The denotation of a formula φ of our propositional language is a set of coalition structures, and we can naturally interpret these as being the desirable or satisfactory coalition structures for a particular player. Thus, instead of writing a hedonic game with dichotomous preferences as a structure $(N, \succeq_1, \ldots, \succeq_n)$, in which we explicitly enumerate preference relations $\succeq_i$, we can instead write $(N, \gamma_1, \ldots, \gamma_n)$, where γ_i is a formula of our propositional language that acts as a specification of the preference relation $\succeq_i$. Because every player i is indifferent between any two partitions that coincide on $\pi(i)$, without loss of generality γ_i involves only propositional variables mentioning i, i.e., it is a formula in the sublanguage L_i of L_N in which only variables in $V_i = \{ij : j \in N \setminus \{i\}\}$ occur.

Intuitively, γ_i represents player i's 'goal' and player i is satisfied if his goal is achieved and unsatisfied if he is not. We refer to a structure $(N, \gamma_1, \ldots, \gamma_n)$ as a *Boolean hedonic game*. Thus, a Boolean hedonic game $(N, \gamma_1, \ldots, \gamma_n)$ represents the (standard) hedonic game $(N, \succeq_1, \ldots, \succeq_n)$ with for each i,

$$\pi \succeq_i \pi' \quad \text{iff} \quad \pi' \models \gamma_i \text{ implies } \pi \models \gamma_i.$$

Observe that, defined thus, the preferences of each player in a hedonic Boolean game are dichotomous.

Often, the use of propositional formulas γ_i gives a 'concise' representation of the preference relation $\succeq_i$, although of course in the worst case the shortest formula γ_i representing $\succeq_i$ may be of size exponential in the number of players. In what follows, we will write $(N, \gamma_1, \ldots, \gamma_n)$, understanding that we are referring to the game $(N, \succeq_1, \ldots, \succeq_n)$ corresponding to this specification.

Example 1 (continued) *The hedonic game with dichotomous preferences in Example 1 is represented by the Boolean hedonic game $(N, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ with $N = \{1, 2, 3, 4\}$ and the players' goals given by:*

$$\begin{array}{ll} \gamma_1 = 123 \vee 124 \vee 134 & \gamma_2 = 213\overline{4} \vee 214\overline{3} \vee 234\overline{1} \\ \gamma_3 = (31 \vee 32) \wedge \neg 34 & \gamma_4 = \neg 423. \end{array}$$

Then, $\pi \models \gamma_i$ if and only if $\pi \in \mathcal{N}_i^+$, for each player i.

Substitution and Deviation

We establish a formal link between substitution in formulas of our language and the possibility of players deviating from their respective coalition in a given partition and joining other coalitions.

Substitution We first introduce some notation and terminology with respect to substitution of formulas for variables in our logic.

For ij a propositional variable in V_N and φ and ψ formulas of L_N, we denote by $\varphi_{ij\leftarrow\psi}$ the *uniform substitution* of variable ij by ψ in φ. If $\vec{ij} = i_1j_1, \ldots, i_kj_k$ is a sequence of k distinct variables in V and $\vec{\psi} = \psi_1, \ldots, \psi_k$ a sequence of k formulas,

$$\varphi_{\vec{ij}\leftarrow\vec{\psi}} \;=\; \varphi_{i_1j_1,\ldots,i_kj_k\leftarrow\psi_1,\ldots,\psi_k}$$

denotes the *simultaneous substitution* of each i_mj_m by ψ_m ($1 \le m \le k$). Thus, e.g., $(ij \vee \neg jk)_{ij,jk\leftarrow jk,ik} = jk \vee \neg ik$. A special case, which recurs frequently in what follows, is if every ψ_i is a Boolean, i.e., if $\psi_1, \ldots, \psi_k \in \{\top, \bot\}$. Sequences $\vec{b} = b_1, \ldots, b_k$ where $b_1, \ldots, b_k \in \{\top, \bot\}$ we will also refer to as *Boolean vectors of length* k. Thus, e.g., $\top, \bot$ is a Boolean vector of length 2 and $(ij \wedge jk \rightarrow ki)_{ij,ki\leftarrow\top,\bot} = \top \wedge jk \rightarrow \bot$.

Characterising Individual Deviations Some of the stability concepts for Boolean hedonic games we consider, e.g., Nash stability, are based on which coalitions an individual player i can join given a partition π. For instance, let partition π be given by $[\,12\,|\,34\,|\,5\,]$. Then, player 1 can join coalition $\{3, 4\}$ but cannot form a coalition with players 4 and 5 by unilaterally deviating from π. We find that the coalitions in $\pi|_{-i}$ can be characterised in our logic. This yields a logical characterisation of when a player i can unilaterally break loose from his coalition, join another one and thereby guarantee that a given formula φ will be satisfied. A particularly interesting case is if φ implies the respective player's goal. We thus gain expressive power with respect to whether a player can *beneficially* deviate from a given partition, a crucial concept. We make this precise in the following two lemmas, where we say, for a given player i, that enumeration $\vec{ij} = ij_1, \ldots, ij_{n-1}$ of V_i and Boolean vector $\vec{b} = b_1, \ldots, b_{n-1}$ *induce* set $B \subseteq N \setminus \{i\}$ if

$$B = \{j_k : ij_k \in V_i \text{ and } b_k = \top\}.$$

Lemma 1 *Let π be a partition, i a player and $B \subseteq N \setminus \{i\}$ induced by enumeration $\vec{ij}$ of V_i and Boolean vector $\vec{b}$. Then,*

$$B \in \pi|_{-i} \;\text{ iff }\; \pi \models \mathit{trans}_{\vec{ij}\leftarrow\vec{b}}.$$

Proof: As $\vec{b}$ and $\vec{ij}$ are fixed throughout the proof, for better readability, we write φ' for $\varphi_{\vec{ij}\leftarrow\vec{b}}$. For the "only if"-direction, assume that $B \in \pi|_{-i}$ and for contradiction that $\pi \not\models \mathit{trans}'$. Observe that $\mathit{trans}' = \bigwedge_{k,l,m} \big(kl' \wedge lm' \rightarrow km'\big)$. Accordingly, there are some mutually distinct k, l, and m such that $\pi \not\models kl' \wedge lm' \rightarrow km'$. It suffices to distinguish three cases:

$$(a) \quad i \notin \{k, l, m\}, \qquad (b) \quad i = k, \qquad (c) \quad i = l.$$

Case (a) cannot occur as we would have $kl' = kl$, $lm' = lm$, $km' = km$, and $kl \wedge lm \rightarrow km$ is a theorem of the system.

If (b), let $l = j_l$ and $m = j_m$. Then $\pi \not\models il' \wedge lm' \rightarrow im'$. It follows that $\pi \models il'$, $\pi \models lm'$, and $\pi \not\models im'$. Observe that in this case $lm' = lm$. Hence, $\pi(l) = \pi(m)$. Also notice that $il', im' \in \{\top, \bot\}$ and, thus, $il' = b_l = \top$ and $im' = b_m = \bot$. Accordingly, $l \in B$ but $m \notin B$. Having assumed that $B \in \pi|_{-i}$, we may conclude that $\pi(l) \neq \pi(m)$ and a contradiction follows.

If (c), let $k = j_k$ and $m = j_m$. We have $\pi \not\models ik' \wedge im' \rightarrow km'$. Thus, $\pi \models ik'$, $\pi \models im'$, and $\pi \not\models km'$. Observe that $km' = km$. Hence, $\pi(k) \neq \pi(m)$. Moreover, $ik', im' \in \{\top, \bot\}$, from which follows that $ik' = b_k = \top$ and $im' = b_m = \top$. Accordingly, both $k, m \in B$. With $B \in \pi|_{-i}$, we obtain $\pi(k) = \pi(m)$, a contradiction.

For the "if"-direction, assume $B \notin \pi|_{-i}$. Then, $B \neq \emptyset$ and let $k = j_k \in B$. Thus, $ik' = b_k = \top$. As $B \notin \pi|_{-i}$, and in particular $B \neq \pi(k) \setminus \{i\}$, there are two possibilities:

(1) there is some $m \neq i$ with $m \in \pi(k)$ and $m \notin B$, or
(2) there is some $m \neq i$ with $m \notin \pi(k)$ and $m \in B$.

If (1), let $m = j_m$ and we have both $\pi(k) = \pi(m)$ and $im' = b_m = \bot$. As $km' = km$, it holds that $\pi \models ik' \wedge km'$ but $\pi \not\models im'$. If (2), also let $m = j_m$. Then, however, we have $\pi(k) \neq \pi(m)$ and $im' = b_m = \top$. As $km' = km$, it holds that $\pi \models ik' \wedge im'$ but $\pi \not\models km'$. In either case $\pi \not\models \mathit{trans}'$. □

The proof of Lemma 2 is by structural induction on φ and relies on similar principles as Lemma 1.

Lemma 2 *Let π be a partition, i a player, and $B \in \pi|_{-i}$ induced by enumeration $\vec{ij}$ of V_i and Boolean vector $\vec{b}$. Then,*

$$\pi[i \rightarrow B] \models \varphi \;\text{ iff }\; \pi \models (\varphi)_{\vec{ij}\leftarrow\vec{b}}.$$

The following example illustrates Lemma 2.

Example 2 *Consider the partition $\pi = [12\,|\,34\,|\,5]$. Then, $\pi|_{-1} = [2\,|\,34\,|\,5]$. Let $\vec{1j} = 12, 13, 14, 15$ be a fixed enumeration of V_1. Also let $\vec{b} = \bot, \top, \top, \bot$ and $\vec{b}' = \bot, \top, \bot, \top$ be Boolean vectors (of length 4). Then,*

$$[\,12\,|\,34\,|\,5\,] \models \mathit{trans}_{12,13,14,15\leftarrow\bot,\top,\top,\bot},$$

which means that 1 can deviate in such a way that he becomes in the same coalition as 3 and 4, but not as 2 or 5. (This may be established by checking all 30 conjuncts of the form $(kl \wedge lm) \rightarrow km$ of trans.) Now, observe that the set induced by $\vec{1j}$ and $\vec{b}$ is $\{3, 4\}$, and thus included in $\pi|_{-1}$. On the other hand, observe that $(13 \wedge 15 \rightarrow 35)_{\vec{1j}\leftarrow\vec{b}_2} = (\top \wedge \top) \rightarrow 35$. Then, $[12|34|5]$ does not satisfy $(\top \wedge \top) \rightarrow 35$ and, hence, neither $\mathit{trans}_{\vec{1j}\leftarrow\vec{b}_2}$. Finally, observe that $\{3, 5\}$, the set induced by $\vec{ij}$ and $\vec{b}'$, is not in $\pi|_{-1}$.

We now introduce the following abbreviation, where $\vec{ij} = ij_1, \ldots, ij_{n-1}$ is assumed to be a fixed enumeration of V_i:

$$\hat{\exists} i\, \varphi \;=\; \bigvee_{\vec{b}\in\{\bot,\top\}^{n-1}} (\varphi \wedge \mathit{trans})_{\vec{ij}\leftarrow\vec{b}}.$$

Thus, $\hat{\exists} i$ can be understood as the operation of forgetting everything about player i (in the sense of (Lin and Reiter 1994)) while taking the transitivity constraint into account. Intuitively, $\hat{\exists} i\, \varphi$ signifies that given partition π player i can deviate to some coalition such that φ is satisfied.

Proposition 2 *Let π be a partition, i a player, and φ a formula of L_N. Then,*

$$\pi \models \hat{\exists} i\, \varphi \quad \textit{iff} \quad \pi[i \to S] \models \varphi \textit{ for some } S \in \pi|_{-i}.$$

Proof: First assume $\pi \models \hat{\exists} i\, \varphi$. Then, $\pi \models (\varphi \wedge \mathit{trans})_{i\vec{\jmath} \leftarrow \vec{b}}$ for some Boolean vector $\vec{b} = b_1, \ldots, b_{n-1}$. Define S as the set induced by $i\vec{\jmath}$ and $\vec{b}$. As $(\varphi \wedge \mathit{trans})_{i\vec{\jmath} \leftarrow \vec{b}} = (\varphi)_{i\vec{\jmath} \leftarrow \vec{b}} \wedge \mathit{trans}_{i\vec{\jmath} \leftarrow \vec{b}}$, by Lemmas 1 and 2, we then obtain $S \in \pi|_{-i}$ and $\pi[i \to S] \models \varphi$, respectively.

For the opposite direction, assume that $\pi[i \to S] \models \varphi$ for some $S \in \pi|_{-i}$. Define $\vec{b} = b_1, \ldots, b_{n-1}$ as the Boolean vector of length $n-1$ such that for every $1 \leq k \leq n-1$,

$$b_k = \begin{cases} \top & \text{if } j \in S \cup \{i\}, \\ \bot & \text{otherwise.} \end{cases}$$

Then, clearly, S is the set induced by $i\vec{\jmath}$ and $\vec{b}$. By Lemmas 1 and 2, it follows that $\pi \models \mathit{trans}_{i\vec{\jmath} \leftarrow \vec{b}}$ and $\pi \models \varphi_{i\vec{\jmath} \leftarrow \vec{b}}$. We may conclude that $\pi \models \hat{\exists} i\, \varphi$. □

Note, however, that the number of Boolean vectors of length k is exponential in k. Accordingly, $\hat{\exists} i\, \varphi$ abbreviates a formula whose size is exponential in the size of φ.

Characterising Group Deviations Besides a single player deviating from its coalition and joining another, multiple players (from possibly different coalitions) could also deviate together and form a coalition of their own. This concept lies at the basis of, e.g., the core stability concept. We characterise group deviations through substitution.

Let $T = \{i_1, \ldots, i_t\}$ be a group of players. Observe that $|V_T| = \binom{n}{2} - \binom{n-t}{2}$ and let $\vec{\imath\jmath}_T$ be a fixed enumeration of V_T. Given $\vec{\imath\jmath}_T$, the we define the *T-separating Boolean vector* $\vec{b}_T$ as the unique Boolean vector of length $\binom{n}{2} - \binom{n-t}{2}$ such that for all $i \in T$ and all $j \in N$,

$$ij_{\vec{\imath\jmath}_T \leftarrow \vec{b}_T} = \begin{cases} \top & \text{if } j \in T, \\ \bot & \text{otherwise.} \end{cases}$$

Intuitively, $\vec{b}_T$ represents the choice of group T to form a coalition of their own. Whenever T is clear from the context we omit the subscript in $\vec{b}_T$ and $\vec{\imath\jmath}_T$. The following characterisation now holds.

Lemma 3 *Let T a group of players, π a partition, $\vec{\imath\jmath}$ a fixed enumeration of V_T, and $\vec{b}_T$ the corresponding T-separating Boolean vector. Then, for every formula $\varphi \in L_N$,*

$$\pi \models (\varphi)_{\vec{\imath\jmath} \leftarrow \vec{b}_T} \quad \textit{iff} \quad \pi[T \to \emptyset] \models \varphi.$$

Characterising Solutions

Our task in this section is to show how the various solution concepts we introduced above can be *characterised* as formulas of our propositional language. Let f be a function mapping each Boolean hedonic game G for N to a formula $f(G)$ of L_N. Given a solution concept θ, we say that f is a *characterisation* of θ if for every Boolean hedonic game G on N and every partition π, we have that π is a solution according to θ for game G if and only if $\pi \models f(G)$. If, furthermore, there exists a polynomial p such that $|f(G)| \leq p(|N|)$, then f is a *polynomial characterisation* of θ.

Once we have a characterisation of θ, we know that there is a one-to-one correspondence between the partitions of N satisfying θ and the models of $f(G)$, i.e., our characterisation is *model-preserving*.[1] Therefore, given a Boolean hedonic game G:

- checking whether there exists a partition satisfying θ in G amounts to checking whether $f(G)$ is satisfiable;
- computing a partition satisfying θ in G amounts to finding a model of $f(G)$;
- computing all partitions satisfying θ in G amounts to finding all models of $f(G)$.

Thus, once we have a characterisation of a solution concept, one may want to use a SAT solver to find (some or all) or to check the existence of partitions that satisfy it. This carries over to *conjunctions* of solution concepts. For instance, if individual rationality is characterised by f_{IR} and envy-freeness by f_{EF}, the there is a one-to-one correspondence between the individual rational envy-free partitions for G and the models of $f_{IR}(G) \wedge f_{EF}(G)$. More generally, these techniques can be used for finding or checking partitions satisfying θ that also have other properties expressible in L_N.

In the remainder of the section we focus on a number of classical solution concepts, and see how they can be characterised in our logic.

Individual Rationality, Perfection, and Optimality Recall that a partition is individually rational if any player is at least as happy in her coalition as being alone, i.e., no player would prefer to leave her coalition to form a singleton coalition. Now we have the following characterisation of individual rationality in our logic.

Proposition 3 *Let $(N, \gamma_1, \ldots, \gamma_n)$ be a Boolean hedonic game, i a player, and π be a partition. Let, furthermore, $i\vec{\jmath}$ be a fixed enumeration of V_i and let $\vec{b} = \vec{\bot} = \bot, \ldots, \bot$ be the Boolean vector of length $n-1$ only containing $\bot$. Then,*

(*i*) *π is acceptable to i iff $\pi \models (\gamma_i)_{i\vec{\jmath} \leftarrow \vec{\bot}} \to \gamma_i$,*

(*ii*) *π is individually rational iff $\pi \models \bigwedge_{i \in N} \big((\gamma_i)_{i\vec{\jmath} \leftarrow \vec{\bot}} \to \gamma_i\big)$.*

Proof: We only give the proof for (i), as (ii) follows as an immediate consequence. For (i), observe that the set induced by $i\vec{\jmath}$ and $\vec{\bot}$ is $\emptyset$ and therefore contained in $\pi|_{-i}$. Then consider the following equivalences,

$$\begin{aligned} \pi \text{ is acceptable to } i \quad & \text{iff} \quad \pi \succeq_i \pi[i \to \emptyset] \\ & \text{iff} \quad \pi[i \to \emptyset] \models \gamma_i \text{ implies } \pi \models \gamma_i \\ & \text{iff} \quad \pi \models (\gamma_i)_{i\vec{\jmath} \leftarrow \vec{\bot}} \text{ implies } \pi \models \gamma_i \\ & \text{iff} \quad \pi \models (\gamma_i)_{i\vec{\jmath} \leftarrow \vec{\bot}} \to \gamma_i, \end{aligned}$$

[1]Note that the fact that the existence of a partition satisfying some solution concept is NP-complete does not imply that there is a model-preserving translation into SAT.

of which the third one holds by virtue of Lemma 2. □

Note that these characterisations are of polynomial size. To illustrate Proposition 3 we consider again Example 1.

Example 1 (continued) *In the game of our example, all partitions are acceptable to player* 1*, whose goal is given by* $\gamma_1 = 123 \vee 124 \vee 134$*. Let* V_1 *be enumerated by* $1\vec{\jmath} = 12, 13, 14$ *and let* $\vec{b} = \bot, \bot, \bot$*. Then,* $(\gamma_2)_{12,13,14 \leftarrow \bot,\bot,\bot}$ *is* $\mathbf{P}$*-equivalent to* $\bot$ *and, hence,* $\pi \models (\gamma_2)_{12,13,14 \leftarrow \bot,\bot,\bot} \rightarrow \gamma_1$ *for all partitions* π*. According to Proposition 3 this signifies that to player* 1 *every partition is acceptable.*

Now consider player 4*, whose goal is given by* $\neg 423$*, i.e., by* $\neg(42 \wedge 43)$*. Let* V_4 *be enumerated by* $41, 42, 43$ *and let* $\vec{b} = \bot, \bot, \bot$*. Then,* $\neg(42 \wedge 43)_{41,42,43 \leftarrow \bot,\bot,\bot} = \neg(\bot \wedge \bot)$*, which is obviously* $\mathbf{P}$*-equivalent to* $\top$*. Hence,*

$$\pi \models \neg(42 \wedge 43)_{41,42,43 \leftarrow \bot,\bot,\bot} \quad \textit{iff} \quad \pi \models \neg(42 \wedge 43),$$

meaning that a partition π *is acceptable to player* 4 *if and only if* π *satisfies his goal.*

Thus, we obtain the following logical characterisation.

Proposition 4 *Let* $(N, \gamma_1, \dots, \gamma_n)$ *be a Boolean hedonic game. Then, a partition* π *is perfect* *iff* $\pi \models \bigwedge_{i \in N} \gamma_i$.

Consequently, a perfect partition exists if and only if the formula $\mathit{trans} \wedge \bigwedge_{i \in N} \gamma_i$ is satisfiable. Finding a social welfare maximising partition reduces thus to finding valuation satisfying a maximum number of formulas $\gamma_i \wedge \mathit{trans}$.

Note that deciding whether a perfect partition on a Boolean hedonic game exists is NP-complete (Peters 2016a), which makes this translation into satisfiability even more meaningful.

Leveraging the same idea of iteratively checking whether a perfect partition for a subset of agents exists, one can compute Pareto optimal solutions for a given game. A subset Ψ of formulas is said to be *maximal trans-consistent* if both

(i) $\Psi \cup \{\mathit{trans}\}$ is consistent, and

(ii) $\Psi' \cup \{\mathit{trans}\}$ is inconsistent for all Ψ' with $\Psi \subsetneq \Psi'$.

We now have the following proposition.

Proposition 5 *A partition* π *of a Boolean hedonic game is Pareto optimal if and only if* $\{\gamma_i : \pi \models \gamma_i\}$ *is a maximal* trans*-consistent subset of* $\{\gamma_1, \dots, \gamma_n\}$.

Algorithms for computing maximal consistent subsets are well-known and may be used to find Pareto optimal partitions (cf., e.g., (Ben-Eliyahu and Dechter 1996; Marquis 2000; Lian and Waaler 2008; Liffiton and Sakallah 2008)).

Nash Stability Recall that a partition π is Nash stable, if no player i wishes to leave his coalition $\pi(i)$ and join another coalition so as to satisfy his goal. Exploiting the results from the previous section, we obtain the following characterisation of this fundamental solution concept.

Proposition 6 *Let* $(N, \gamma_1, \dots, \gamma_n)$ *be a Boolean hedonic game and* π *a partition. Then,*

$$\pi \text{ is Nash stable } \textit{iff } \pi \models \bigwedge_{i \in N} \big((\hat{\exists} i\, \gamma_i) \rightarrow \gamma_i\big).$$

Proof: Consider an arbitrary player i and observe that following equivalences hold.

π is Nash stable
iff for all i and $S \in \pi|_{-i}$, $\pi \succeq_i \pi[i \rightarrow S]$
iff for all i and $S \in \pi|_{-i}$, if $\pi[i \rightarrow S] \models \gamma_i$, then $\pi \models \gamma_i$
iff for all i, if $\pi[i \rightarrow S] \models \gamma_i$ for an $S \in \pi|_{-i}$, then $\pi \models \gamma_i$
iff for all i, if $\pi \models \hat{\exists} i\, \gamma_i$ then $\pi \models \gamma_i$
iff for all i, $\pi \models (\hat{\exists} i\, \gamma_i) \rightarrow \gamma_i$
iff $\pi \models \bigwedge_{i \in N} \big((\hat{\exists} i\, \gamma_i) \rightarrow \gamma_i\big)$.

The fourth equivalence holds by virtue of Proposition 2. The third one is a standard law of logic: merely observe that whether $\pi \models \gamma_i$ is not dependent on S. □

Our running example illustrates this result.

Example 1 (continued) *Consider again the game of Example 1. Partition* [123|4] *satisfies each player's goal and, consequently, is Nash stable. We also have that* $[123|4] \models \gamma_1 \wedge \gamma_2 \wedge \gamma_3 \wedge \gamma_4$ *and, thus,*

$$[123|4] \models \bigwedge_{i \in N} \big((\hat{\exists} i\, \gamma_i) \rightarrow \gamma_i\big).$$

Now recall that for partition $\pi = [1|23|4]$ *player* 2*'s goal is not satisfied and that she cannot deviate and join another coalition to make this happen. In this case,* $\pi|_{-2} = \{\{1\}, \{3\}, \{4\}\}$*. Moreover,* $\pi[2 \rightarrow \{1\}] = [12|3|4]$*,* $\pi[2 \rightarrow \{3\}] = [1|23|4]$*, and* $\pi[2 \rightarrow \{4\}] = [1|3|24]$*. Since,* $[12|3|4] \not\models \gamma_2$*,* $[1|23|4] \not\models \gamma_2$*, and* $[1|3|24] \not\models \gamma_2$*, it follows that* $\pi \not\models \hat{\exists} 2\, \gamma_2$*. Hence,* $\pi \models (\hat{\exists} 2\, \gamma_2) \rightarrow \gamma_2$*. Player* 1*, however, could deviate from* π_2 *and join* $\{2, 3\}$ *and thus have his goal satisfied. Thus,* π *is not Nash stable. Now observe that* $\{2, 3\} \in \pi|_{-1}$ *and that* $\pi[1 \rightarrow \{2, 3\}] = [123|4]$*. Moreover,* $[123|4] \models \gamma_1$*. As thus* $\pi \models \hat{\exists} 1\, \gamma_1$*, also* $\pi \not\models (\hat{\exists} 1\, \gamma_1) \rightarrow \gamma_1$*. We may conclude that*

$$[1|23|4] \not\models \bigwedge_{i \in N} \big((\hat{\exists} i\, \gamma_i) \rightarrow \gamma_i\big).$$

Nash stable partitions are not guaranteed to exist in Boolean hedonic games. The two-player game $(\{1, 2\}, 12, \neg 21)$ witnesses this fact, as can easily be appreciated. The translation into a SAT instance gives us a way to compute all Nash stable partitions of a given Boolean hedonic game. Recall, however, that the size of $\hat{\exists} i\, \gamma_i$ is generally exponential in the size of γ_i. Of course, one may wonder why this is useful to express Nash stability as an exponentially large SAT instance, since the existence of a Nash stable partition is "only" NP-complete (Peters 2016a). However, we stress that our translation is *model-preserving*, which means that it is particularly useful if we want to use Nash stability in conjunction with other concepts; moreover, in many practical cases, the translation will remain of reasonable size.

Core and Strict Core Stability Core and strict core stability relate to group deviations much in the same way as

Nash stability relates to individual deviations. Having characterised group deviations in the previous section, we find that Lemma 3 yields a straightforward characterisation in our logic of a specific group blocking or weakly blocking a given partition.

Proposition 7 *Let $(N, \gamma_1, \ldots, \gamma_n)$ be a Boolean hedonic game and T a group of players, and π be a partition. Let, furthermore, $\vec{ij}$ a fixed enumeration of V_T and $\vec{b}_T$* the corresponding T-separating Boolean vector. *Then,*

(i) *T blocks π iff* $\pi \models \bigwedge_{i \in T} \big(\neg\gamma_i \wedge (\gamma_i)_{\vec{ij} \leftarrow \vec{b}_T}\big)$,

(ii) *T weakly blocks π iff*

$$\pi \models \bigwedge_{j \in T} \big(\gamma_j \to (\gamma_j)_{\vec{ij} \leftarrow \vec{b}_T}\big) \wedge \bigvee_{i \in T} \big(\neg\gamma_i \wedge (\gamma_i)_{\vec{ij} \leftarrow \vec{b}_T}\big).$$

Proof: We give the proof for (i), as the one for (ii) runs along analogous lines. Consider the following equivalences:

T blocks π iff for all $i \in T$: $\pi[T \to \emptyset] \succ_i \pi$
iff for all $i \in T$: $\pi[T \to \emptyset] \models \gamma_i$ and $\pi \not\models \gamma_i$
iff for all $i \in T$: $\pi \models (\gamma_i)_{\vec{ij} \leftarrow \vec{b}}$ and $\pi \not\models \gamma_i$
iff $\pi \models \bigwedge_{i \in T} \big(\neg\gamma_i \wedge (\gamma_i)_{\vec{ij} \leftarrow \vec{b}}\big)$.

The third equivalence holds by virtue of Lemma 3. □

Observe that the size of $\bigwedge_{i \in T} \big(\neg\gamma_i \wedge (\gamma_i)_{\vec{ij} \leftarrow \vec{b}}\big)$ is polynomial in $\sum_{i \in T} |\gamma_i|$ and, hence, a partition π being blocking by a group T of players can be polynomially characterised.

As a corollary of Proposition 7 and the de Morgan laws, we obtain the following characterisations of a partition being core stable and of a partition being strict core stable. The characterisations, however, involve a conjunctions over all groups of players and as such is not polynomial.

Corollary 1 *Let $(N, \gamma_1, \ldots, \gamma_n)$ be a Boolean hedonic game and π be a partition. Let for each coalition T, $\vec{ij}$ be an enumeration of V_T and $\vec{b}$ the corresponding T-separating Boolean vector. Then,*

(i) *π is core stable iff* $\pi \models \bigwedge_{T \subseteq N} \bigvee_{i \in T} \big((\gamma_i)_{\vec{ij} \leftarrow \vec{b}} \to \gamma_i\big)$,

(ii) *Then, π is strict core stable iff*

$$\pi \models \bigwedge_{T \subseteq N} \Big(\bigvee_{j \in T} (\gamma_j \wedge \neg(\gamma_j)_{\vec{ij} \leftarrow \vec{b}}) \vee \bigwedge_{i \in T} \big((\gamma_i)_{\vec{ij} \leftarrow \vec{b}} \to \gamma_i\big)\Big).$$

Although core stable coalition structures are not guaranteed to exist in general hedonic games, if preferences are dichotomous we have the following positive result.

Proposition 8 *For every Boolean hedonic game, a core stable coalition structure is guaranteed to exist.*

Proof: We initialise N' to N and partition π to $\{\emptyset\}$. We find a maximal subset of $S \subset N'$ for which all players are in an approved coalition that satisfies their formulas. We modify π to $\pi \cup \{S\}$ and N' to $N' \setminus S$. The procedure is repeated until no such maximal subset S exists. If $N' \neq \emptyset$, then π is set to $\pi \cup \{\{i\} : i \in N'\}$.

We now argue that π is core stable. We note that each player who was in some subset S will never be part of a blocking coalition. If N' was non-empty in the last iteration, then no subset of players in N' can form a deviating coalition among themselves. □

By contrast, a strict core stable partition is not guaranteed to exist. To see this consider the three-player Boolean hedonic game $(\{1, 2, 3\}, 12, 21 \vee 23, 32)$: each of the five possible partitions is weakly blocked by either $\{1, 2\}$ or $\{2, 3\}$.

The characterisation of strict core stability is not of polynomial size, but it is highly unlikely that such a characterisation exists, since deciding whether there exists a strict core stable partition is Σ_2^{P}-complete (Peters 2016b).

Envy-freeness Recall that a partition is envy-free if no player would strictly prefer to exchange places with another player. Observe that for the trivial partitions $\pi^0 = [1 \,|\, \cdots \,|\, n]$ and $\pi^1 = [1, \ldots, n]$, we have $\pi^0[i \leftrightarrows j] = \pi^0$ and $\pi^1[i \leftrightarrows j] = \pi^1$ for all players i and j. Accordingly π^0 and π^1 are envy-free. Envy-free partitions are thus guaranteed to exist in our setting. The following lemma allows us to derive a polynomial characterisation of envy-freeness.

Lemma 4 *Let $(N, \gamma_1, \ldots, \gamma_n)$ be a Boolean hedonic game and i and j players in N, and φ a formula in L_N. Fix, furthermore, an enumeration $k_1, \ldots, k_{n-2}$ of $N \setminus \{i, j\}$ and let $i\vec{k} = ik_1, \ldots, ik_{n-2}$ and $j\vec{k} = jk_1, \ldots, jk_{n-2}$ enumerate $V_i \setminus \{ij\}$ and $V_j \setminus \{ji\}$, respectively. Then,*

$$\pi \models \varphi_{i\vec{k}, j\vec{k} \leftarrow j\vec{k}, i\vec{k}} \quad \textit{iff} \quad \pi[i \leftrightarrows j] \models \varphi.$$

Proof: With $i\vec{k}$ and $j\vec{k}$ being fixed we write φ' for $\varphi_{i\vec{k}, j\vec{k} \leftarrow j\vec{k}, i\vec{k}}$. The proof is then by a structural induction on φ, of which only the basis is interesting.

Let $\varphi = lm$. There are three possibilities: (a) $lm = ij$, (b) $lm \in (V_i \cup V_j) \setminus \{ij\}$, and (c) $lm \notin V_i \cup V_j$. If (a), we have that $lm' = ij' = ij = lm$. Now, either $\pi(i) = \pi(j)$ or $\pi(i) \neq \pi(j)$. If the former, $\pi[i \leftrightarrows j] = \pi$ as well as both $\pi \models ij'$ and $\pi[i \leftrightarrows j] \models ij$. If the latter, however, it can easily be seen that both $\pi \not\models ij'$ and $\pi[i \leftrightarrows j] \not\models ij$.

For case (b), we may assume without loss of generality that $lm = ik$ for some $k \neq j$. Then, $ik' = jk$. In case $\pi(i) = \pi(j)$, obviously, $\pi = \pi[i \leftrightarrows j]$ as well as $k \in \pi(i)$ if and only if $k \in \pi(j)$. Hence, $\pi \models ik'$ if and only if $\pi[i \leftrightarrows j] \models ik$. So, assume $\pi(i) \neq \pi(j)$. Now, either (i) $k \in \pi(i)$ and $k \notin \pi(j)$, (ii) $k \notin \pi(k)$ and $k \in \pi(j)$, or (iii) $k \notin \pi(i)$ and $k \notin \pi(j)$. If (i), $\pi \models ik'$ as well as $\pi[i \leftrightarrows j] \models jk$. In cases (ii) and (iii), we have $\pi \not\models ik'$ and $\pi[i \leftrightarrows j] \not\models jk$.

Finally, if (c), then $lm' = lm$. As $l, m \notin \{i, j\}$, it is easily seen that $\pi \models lm'$ if and only if $\pi[i \leftrightarrows j] \models lm$. □

We are now in a position to state the following result.

Proposition 9 *Let $(N, \gamma_1, \ldots, \gamma_n)$ be a Boolean hedonic game. Furthermore, for every two players, i and j, and enumeration $k_1, \ldots, k_{n-2}$ of $N \setminus \{i, j\}$, let $i\vec{k} = ik_1, \ldots, ik_{n-2}$ and $j\vec{k} = jk_1, \ldots, jk_{n-2}$ enumerate $V_i \setminus \{ij\}$ and $V_j \setminus \{ij\}$, respectively. Then,*

$$\pi \textit{ is envy-free} \quad \textit{iff} \quad \pi \models \bigwedge_{i,j \in N} \big((\gamma_i)_{i\vec{k}, j\vec{k} \leftarrow j\vec{k}, i\vec{k}} \to \gamma_i\big).$$

Proof: By Lemma 4, the following equivalences hold:

$$\begin{array}{ll} & \pi \text{ is envy-free} \\ \text{iff} & \text{for all } i,j \in N\text{: } \pi \succeq_i \pi[i \leftrightarrows j] \\ \text{iff} & \text{for all } i,j \in N\text{: } \pi[i \leftrightarrows j] \models \gamma_i \text{ implies } \pi \models \gamma_i \\ \text{iff} & \text{for all } i,j \in N\text{: } \pi \models (\gamma_i)_{i\vec{k},j\vec{k} \leftarrow j\vec{k},i\vec{k}} \text{ implies } \pi \models \gamma_i \\ \text{iff} & \text{for all } i,j \in N\text{: } \pi \models (\gamma_i)_{i\vec{k},j\vec{k} \leftarrow j\vec{k},i\vec{k}} \to \gamma_i \\ \text{iff} & \pi \models \bigwedge_{i,j \in N} \big((\gamma_i)_{i\vec{k},j\vec{k} \leftarrow j\vec{k},i\vec{k}} \to \gamma_i \big). \end{array}$$

This concludes the proof. □

Observe that the size of $\bigwedge_{i,j\in N} \big((\gamma_i)_{i\vec{k},j\vec{k} \leftarrow j\vec{k},i\vec{k}} \to \gamma_i \big)$ is polynomial in $\sum_{i \in T} |\gamma_i|$, hence we get a polynomial characterization of envy-freeness. Note that an envy-free partition always exists (e.g., the partition where all agents are together is envy-free, or the partition where they are all isolated), but once again, the fact that the translation is model-preserving allows us to compute *all* envy-free partitions.

Example 1 (continued) *Recall that* $\gamma_3 = (31 \vee 32) \wedge \neg 34$ *and that player* 3 *envies player* 4 *if partition* $\pi' = [1|24|3]$ *obtains. To see how this is reflected by Proposition 9, let* $31, 32$ *and* $41, 42$ *enumerate* $V_3 \setminus \{34\}$ *and* $V_4 \setminus \{43\}$*, respectively. Then,*

$$((31 \vee 32) \wedge \neg 34)_{31,32,41,42 \leftarrow 41,42,31,32} = (41 \vee 42) \wedge \neg 34.$$

Now, both $\pi' \models (41 \vee 42) \wedge \neg 34$ *and* $\pi' \not\models (31 \vee 32) \wedge \neg 34$*, and, hence,* $\pi' \not\models (\gamma_3)_{34,31,32 \leftarrow 43,41,42} \to \gamma_3$.

After some simplifications, the formulas expressing that players 1, 2, 3, 4 *are not envious are*

for 1: $234 \to 1234$ *for* 2: $\neg 134\bar{2}$
for 3: $\neg 12\bar{3}\bar{4} \wedge \neg 41\bar{3} \wedge \neg 42\bar{3}$ *for* 4: $\neg 423 \vee 1234$.

Therefore, π *is envy-free if and only if it satisfies*

$$1234 \vee (\neg 234 \wedge \neg 134 \wedge \neg 12\bar{3}\bar{4} \wedge \neg 41\bar{3} \wedge \neg 42\bar{3})$$

or, equivalently, $1234 \vee (\neg 24 \wedge \neg 14 \wedge \neg 12\bar{3}\bar{4})$. *This formula (that we call* EF*) is satisfied by the partitions* [1234], [123|4], [1|2|34], [13|2|4], [1|23|4] *and* [1|2|3|4]. *Now, we may want to require envy-freeness in addition with another property. For instance, assume there should be exactly two coalitions, which is expressed by the polynomial-size formula* $\psi = \bigvee_{1 \le i < j \le 4} \neg ij \wedge \bigwedge_{1 \le i<j<k \le 4} \neg(i\bar{j}\bar{k} \wedge j\bar{i}\bar{k} \wedge k\bar{i}\bar{j})$. $EF \wedge \psi$ *has a single model corresponding to* [4|123].

Related Work and Conclusions

Our motivation and approach is strongly reminiscent of the setting of Boolean games in the context of non-cooperative game theory (Harrenstein et al. 2001). A major difference with Boolean games and propositional hedonic games is that in Boolean games, players have preferences over outcomes, where an outcome is a truth assignment to outcome variables, and each outcome variable is controlled by a specific player. This control assignment function, which is a central notion in Boolean games, has no counterpart here, where the outcome is a partition of the players. However, there are technical similarities with and conceptual connections to Boolean games, especially when characterising solution concepts. For instance, the characterisation of Nash stable partitions by propositional formulas (Section 4) is similar to the characterisation of Nash equilibria by propositional formulas in Boolean games as by Bonzon et al. (2009). The basic Boolean games model of Harrenstein et al. (2001) was adapted to the setting of cooperative games by Dunne et al. (2008); there, however, the logic used to specific player's goals was not intended for specifying desirable coalition structures, as we have done in the present paper.

Our work also shares some common ground with the work of Bonzon, Lagasquie-Schiex, and Lang (2012), who study the formation of efficient coalitions in Boolean games, i.e., coalitions whose joint abilities allow their members to jointly achieve their goals. Our work also bears some resemblance to the work of Elkind and Wooldridge (2009), who were interested in using logic as a foundation upon which to build a compact representation scheme for hedonic games; more precisely, their work made use of weighted Boolean formulas, and was inspired by the *marginal contribution nets* representation for cooperative games proposed by Ieong and Shoham (2005). Hedonic Coalition Nets are intended for modelling arbitrary hedonic games, rather than hedonic games with dichotomous preferences. The "logic" component of HC-nets uses Boolean conditions on rules, but the logical aspects of HC-nets are not developed nearly so deeply as our formalism. Moreover, the focus of Elkind and Wooldridge (2009) is more on complexity issues than in finding exact characterisations for solution concepts.

Our characterisations of solution concepts enable the use of off-the-shelf SAT solvers to compute partitions satisfying a solution concept or a logical combination of solution concepts. The complexity of finding and deciding the existence of such partitions in Boolean hedonic games has been studied recently by Peters (2016a; 2016b). Building upon a first version of our paper (Aziz et al. 2015), he shows that all problems related to Boolean hedonic games (except individual stability, which we did not study here) are computationally hard. This supplements our results and makes our characterisations more interesting.

There are at least two directions in which our work might be further developed. First, we could think of relaxing our restriction to dichotomous preferences and study more general hedonic games with compact logical representations, such as prioritised goals or weighted goals, and derive exact characterisations of solution concepts. For some of the cruder extensions our results extend naturally and straightforwardly. For the more sophisticated settings more research is required, which falls beyond the scope of this paper.

Second, our restriction to *hedonic* preferences could be relaxed, so that players have preferences that depend not only on the coalition to which they belong. This would pave the way to a more general *logic of coalition structures*. Solution concepts, once generalised, can hopefully be characterised.

A third topic of future research would be the characterisation of classes of hedonic games in our logic. As mentioned above, various classes of hedonic games that allow for a concise representation have been proposed in the literature (see, e.g., (Aziz and Savani 2016; Peters and Elkind 2015)). It would be interesting to see whether these classes can also be polynomially characterised in our logic.

Acknowledgments We thank Dominik Peters and anonymous reviewers for comments. Haris Aziz is supported by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program. Jérôme Lang is supported by the ANR project CoCoRICo-CoDec. Paul Harrenstein and Michael Wooldridge are supported by the ERC under Advanced Investigator Grant 291528 ("RACE").

References

Aziz, H., and Savani, R. 2016. Hedonic games. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*. Cambridge University Press. chapter 15.

Aziz, H.; Harrenstein, P.; Lang, J.; and Wooldridge, M. 2015. Boolean hedonic games. *arXiv preprint* arXiv:1509.07062. Manuscript. Also presented at the *Eleventh Conference on Logic and the Foundations of Game and Decision Theory (LOFT-2014)*, Bergen, Norway.

Aziz, H.; Brandt, F.; and Harrenstein, P. 2013. Pareto optimality in coalition formation. *Games and Economic Behavior* 82:562–581.

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

Ben-Eliyahu, R., and Dechter, R. 1996. Default reasoning using classical logic. *Artificial Intelligence* 84(1):113–150.

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

Bogomolnaia, A., and Moulin, H. 2004. Random matching under dichotomous preferences. *Econometrica* 72(1):257–279.

Bogomolnaia, A.; Moulin, H.; and Stong, R. 2005. Collective choice under dichotomous preferences. *Journal of Economic Theory* 122(2):165–184.

Bonzon, E.; Lagasquie-Schiex, M.-C.; Lang, J.; and Zanuttini, B. 2009. Compact preference representation and Boolean games. *Autonomous Agents and Multi-Agent Systems* 18(1):1–35.

Bonzon, E.; Lagasquie-Schiex, M.-C.; and Lang, J. 2012. Effectivity functions and efficient coalitions in Boolean games. *Synthese* 187(1):73–103.

Bouveret, S., and Lang, J. 2008. Efficiency and envy-freeness in fair division of indivisible goods: Logical representation and complexity. *Journal of AI Research* 32:525–564.

Brams, S. J., and Fishburn, P. C. 2007. *Approval Voting*. Springer, 2nd edition.

Cechlárová, K., and Hajduková, J. 2004. Stable partitions with $\mathcal{W}$-preferences. *Discrete Applied Mathematics* 138(3):333–347.

Cechlárová, K. 2008. Stable partition problem. In *Encyclopedia of Algorithms*. Springer. 885–888.

Chalkiadakis, G.; Elkind, E.; and Wooldridge, M. 2011. *Computational Aspects of Cooperative Game Theory*. Morgan-Claypool.

Drèze, J. H., and Greenberg, J. 1980. Hedonic coalitions: Optimality and stability. *Econometrica* 48(4):987–1003.

Dunne, P. E.; Kraus, S.; van der Hoek, W.; and Wooldridge, M. 2008. Cooperative Boolean games. In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2008)*, 1015–1022.

Elkind, E., and Wooldridge, M. 2009. Hedonic coalition nets. In *Proceedings of the Eighth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2009)*, 417–424.

Hajduková, J. 2006. Coalition formation games: A survey. *International Game Theory Review* 8(4):613–641.

Harrenstein, P.; van der Hoek, W.; Meyer, J.-J.; and Witteveen, C. 2001. Boolean games. In van Benthem, J., ed., *Proceedings of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, 287–298.

Ieong, S., and Shoham, Y. 2005. Marginal contribution nets: A compact representation scheme for coalitional games. In *Proceedings of the Sixth ACM Conference on Electronic Commerce (EC'05)*, 193–202.

Konieczny, S., and Pino-Pérez, R. 2002. Merging information under constraints: a logical framework. *Journal of Logic and Computation* 12(5):773–808.

Lian, E. H., and Waaler, A. 2008. Computing default extensions by reductions on O^R. In *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning*, 496–506.

Liffiton, M. H., and Sakallah, K. A. 2008. Algorithms for computing minimal unsatisfiable subsets of constraints. *Journal of Automated Reasoning* 40(1):1–33.

Lin, F., and Reiter, R. 1994. Forget it! In *Working Notes of AAAI Fall Symposium on Relevance*, 154–159.

Marquis, P. 2000. Consequence finding algorithms. In Gabbay, D. M., and Smets, P., eds., *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 5. Springer. 41–145.

Peters, D., and Elkind, E. 2015. Simple causes of complexity in hedonic games. In Yang, Q., and Wooldridge, M., eds., *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI 2015)*, 617–623. AAAI Press.

Peters, D. 2016a. Computational complexity of hedonic games with dichotomous preferences. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16)*.

Peters, D. 2016b. Σ_2^{P}-complete problems on hedonic games. *arXiv preprint* arXiv:1509.02333. Manuscript.