Cooperative Boolean Games

Paul E. Dunne* Wiebe van der Hoek* Sarit Kraus+ Michael Wooldridge+

*Department of Computer Science
University of Liverpool
Liverpool l69 3BX, UK

+Department of Computer Science
Bar-Ilan University
Ramat Gan, Israel 52900

ABSTRACT

We present and formally investigate *Cooperative Boolean Games*, a new, natural family of coalitional games that are both compact and expressive. In such a game, an agent's primary aim is to achieve its individual goal, which is represented as a propositional logic formula over some set of Boolean variables. Each agent is assumed to exercise unique control over some subset of the overall set of Boolean variables, and the set of valuations for these variables corresponds to the set of actions the agent can take. However, the actions available to an agent are assumed to have some cost, and an agent's secondary aim is to minimise its costs. Typically, an agent must cooperate with others because it does not have sufficient control to ensure its goal is satisfied. However, the desire to minimise costs leads to preferences over possible coalitions, and hence to strategic behaviour. Following an introduction to the formal framework of Cooperative Boolean Games, we investigate solution concepts of the core and stable sets for them. In each case, we characterise the complexity of the associated solution concept, and discuss the surrounding issues. Finally, we present a bargaining protocol for cooperation in Boolean games, and characterise the strategies in equilibrium for this protocol.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; I.2.4 [**Knowledge representation formalisms and methods**]

General Terms

Theory

Keywords

cooperative games, logic, games, complexity, boolean games

1. INTRODUCTION

Cooperative games [11, pp.255–312] provide an important theoretical foundation for cooperation in multi-agent systems [15, 14]. One of the key issues in the application of cooperative games in multi-agent systems is that of *compactly representing them*. Naive representations for cooperative games are exponentially large in the number of agents, and so current literature considers the merits of various compact representation schemes for them, and the implications of these representations for the complexity of computing cooperative solution concepts such as the core [4, 8, 3, 16, 17].

In this paper, we present and formally investigate a new, natural family of compact and expressive coalitional games. In a *Cooperative Boolean Game* (CBG), each agent desires to accomplish a personal goal, which is represented as a propositional logic formula over some set of Boolean variables. Intuitively, the goal formula defines the set of joint actions that the agent would like to see carried out (i.e., the satisfying assignments to the goal formula). Each agent is assumed to exercise unique control over some subset of the overall set of Boolean variables; the set of valuations to these variables corresponds to the set of actions the agent can take. However, the actions available to an agent are assumed to have some cost, and an agent's secondary goal is to minimise its costs. Typically, an agent must cooperate with others because it does not have sufficient control to ensure its goal is satisfied. However, the desire to minimise costs leads to preferences over possible joint actions, and hence over coalitions, leading to strategic behaviour.

CBGs derive in part from non-cooperative Boolean games as proposed by Harrenstein *et al* [6, 5] and further developed by Bonzon *et al* [2, 1]. In non-cooperative Boolean games, as in CBGs, agents have goals represented by propositional formulae, and control some set of variables, but there is no cost element, and strategic concerns arise largely from considerations about how other agents will try to satisfy their goals. CBGs are also descended from Qualitative Coalitional Games (QCGs) [16] and Coalitional Resource Games (CRGs) [17]. In a QCG, agents have goals to achieve, but these goals are essentially restricted to be disjunctive, and there is no cost element to achieving goals. CRGs are a generalisation of QCGs in which the accomplishment of goals is assumed to require the consumption of resources of various kinds; the key issues studied in [17] relate to efficient resource usage, and potential conflicts between coalitions with respect to resource bounds.

The remainder of this paper is structured as follows. After an introduction to the formal framework of Cooperative Boolean Games, we investigate solution concepts for them: the core, and stable sets.. In each case, we characterise the complexity of the associated solution concept, and discuss the surrounding issues. We then give a negotiation protocol for cooperation in Boolean games, and characterise the equilibrium strategies for this protocol. *We assume some familiarity with cooperative games [11], propositional logic, and computational complexity [12].*

2. COOPERATIVE BOOLEAN GAMES

Propositional Logic: Throughout the paper, we make use of classical propositional logic, and for completeness, we thus begin by recalling the technical framework of this logic. Let $\Phi = \{p, q, \ldots\}$

Cite as: Cooperative Boolean Games, P. E. Dunne, W. van der Hoek, S. Kraus, and M. Wooldridge, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 1015-1022.

be a (finite, fixed, non-empty) vocabulary of Boolean variables, and let $\mathcal{L}$ denote the set of (well-formed) formulae of propositional logic over Φ, constructed using the conventional Boolean operators ("$\wedge$", "$\vee$", "$\rightarrow$", "$\leftrightarrow$", and "$\neg$"), as well as the truth constants "$\top$" (for truth) and "$\bot$" (for falsity). We assume a conventional semantic consequence relation "$\models$" for propositional logic. A subset ξ of Φ is a *valuation*, and we write $\xi \models \varphi$ to mean that φ is true under, or satisfied by valuation ξ. Where $\Delta \subseteq \mathcal{L}$, we write $\Delta \models \varphi$ to mean that φ is a logical consequence of Δ. We write $\models \varphi$ if φ is a tautology. We denote the fact that formulae $\varphi, \psi \in \mathcal{L}$ are logically equivalent by $\varphi \Leftrightarrow \psi$; thus $\varphi \Leftrightarrow \psi$ means that $\models \varphi \leftrightarrow \psi$. Note that "$\Leftrightarrow$" is a meta-language relation symbol, which should not be confused with the object-language bi-conditional operator "$\leftrightarrow$". If $\varphi \in \mathcal{L}$, then we let $[\![\varphi]\!]$ be the set of valuations that satisfy φ, i.e., $[\![\varphi]\!] = \{\xi : \xi \subseteq \Phi \ \& \ \xi \models \varphi\}$.

Agents, Goals, and Controlled Variables: The games we consider are populated by a set $A = \{1, \ldots, n\}$ of agents. A *coalition*, typically denoted by C, is simply a (sub)set of agents, $C \subseteq A$. Each agent is assumed to have a *goal*, characterised by an $\mathcal{L}$-formula: we write γ_i to denote the goal of agent $i \in A$. Each agent $i \in A$ *controls* a (possibly empty) subset Φ_i of the overall set of Boolean variables (cf. [7]). By "control", we mean that i has the unique ability within the game to set the value (either $\top$ or $\bot$) of each variable $p \in \Phi_i$. We will require that $\Phi_1, \ldots, \Phi_n$ forms a partition of Φ (i.e., every variable is controlled by some agent, no variable is controlled by more than one agent). Where $C \subseteq A$, we denote by Φ_C the set of variables under the control of some member of C, i.e., $\Phi_C = \bigcup_{i \in C} \Phi_i$. Conversely, given a valuation $\xi \subseteq \Phi$, we denote by $A(\xi)$ the agents controlling variables in ξ, i.e., $A(\xi) = \{i : \exists v \in \xi \text{ s.t. } v \in \Phi_i\}$. Let $contrib(\xi)$ denote the set of agents that incur some cost in ξ: $contrib(\xi) = \{i : \exists v \in \xi \text{ s.t. } v \in \Phi_i \ \& \ c_i(v) > 0\}$; of course, $contrib(\xi) \subseteq A(\xi)$. If a valuation ξ_2 is the same as a valuation ξ_1 except at most in the value of variables controlled by C, (i.e., $(\xi_1 \setminus \xi_2) \cup (\xi_2 \setminus \xi_1) \subseteq \Phi_C$) then we write $\xi_2 = \xi_1 \mod C$. We write $\xi \subseteq_C \xi'$ ($\xi \subset_C \xi'$) if $\xi \cap \Phi_C \subseteq \xi' \cap \Phi_C$ ($\xi \cap \Phi_C \subset \xi' \cap \Phi_C$) and for all $j \notin C$, $\xi \cap \Phi_j = \xi' \cap \Phi_j$. We say that ξ is a C-minimal for φ if $\xi \models \varphi$ and no $\xi' \subset_C \xi$, $\xi' \models \varphi$. Given ξ, we define the *beneficiaries* of ξ as $ben(\xi) = \{i \in A : \xi \models \gamma_i\}$.

Costs: Intuitively, setting a variable $p \in \Phi$ to be $\top$ can be thought of as "performing the action p", while setting this variable to be $\bot$ can be thought of as "doing nothing". Since action (as opposed to inaction) typically incurs some cost, we introduce a *cost function* $c : \Phi \rightarrow \mathbb{R}_+$, so that $c(p)$ denotes the cost of performing the action p (i.e., making p true).

Cooperative Boolean Games: Collecting these components together, a *cooperative Boolean game*, G, is a $(2n+3)$-tuple:

$$G = \langle A, \Phi, c, \gamma_1, \ldots, \gamma_n, \Phi_1, \ldots, \Phi_n \rangle,$$

where $A = \{1, \ldots, n\}$ is a set of agents, $\Phi = \{p, q, \ldots\}$ is a finite set of Boolean variables, $c : \Phi \rightarrow \mathbb{R}_+$ is a cost function, $\gamma_i \in \mathcal{L}$ is the goal of agent $i \in A$, and $\Phi_1, \ldots, \Phi_n$ is a partition of Φ over n, with the intended interpretation that Φ_i is the set of Boolean variables under the unique control of $i \in A$.

Utilities and Preferences: With a slight abuse of notation, we let $c_i(\xi)$ denote the cost to agent $i \in A$ of valuation $\xi \subseteq \Phi$, that is,

$$c_i(\xi) = \sum_{v \in (\xi \cap \Phi_i)} c(v).$$

For convenience, we let μ denote the total cost of *all* variables:

$$\mu = \sum_{v \in \Phi} c(v).$$

The *utility* to agent i of a valuation ξ, denoted $u_i(\xi)$, is defined as:

$$u_i(\xi) = \begin{cases} 1 + \mu - c_i(\xi) & \text{if } \xi \models \gamma_i \\ -c_i(\xi) & \text{otherwise.} \end{cases}$$

The utility function $u_i(\cdot)$ leads naturally to a preference order $\succeq_i$ over valuations:

$$\xi_1 \succeq_i \xi_2 \qquad \text{iff} \qquad u_i(\xi_1) \geq u_i(\xi_2).$$

As usual, we write $\succ_i$ for the corresponding strict preference order. This definition has the following properties:

- an agent prefers all valuations that satisfy its goal over all those that do not satisfy it;
- between two valuations that satisfy its goal, an agent prefers the one that minimises its costs; and
- between two valuations that *do not* satisfy its goal, an agent prefers the one that minimises its costs.

We write $\succ_C$ to mean $\succ_i$ for all $i \in C$. Given this framework, we can describe the "game" that agents play, as follows. An agent's primary objective is, first, to achieve its goal; its secondary objective is to minimise costs. Thus, if the only way an agent can achieve its goal is by making all its variables true, (hence incurring maximum cost to itself), then an agent would prefer to do this rather than not achieve its goal. (This even holds in the extreme case that $\Phi_i = \Phi$ and $\gamma_i = \bigwedge_{p \in \Phi} p$. However, if there are *multiple* ways of achieving its goal, then an agent prefers those that *minimise* costs. The *worst* outcome for agent i is that it doesn't get its goal satisfied, but makes all it variables true, yielding a utility of $-c_i(\Phi_i)$. The *best* outcome for an agent is that it has its goal satisfied without having to make any of its variables true, yielding a utility of $\mu + 1$.

It will not generally be the case that a given agent i will be able to satisfy its goals in isolation: if $\gamma_i = p \wedge q$ and $\Phi_i = \{p\}$, then i will need help if it is to achieve its goal. Alternatively, it may be that two agent's can achieve their goals independently, but by cooperating, they can reduce their respective costs. In sum, agents will cooperate when a cooperative solution is preferable to the alternatives, either because it reduces costs or makes it possible for an agent to achieve a goal that it would not otherwise be able to achieve. Of course, this does not say anything of *how* agents will choose to cooperate – *which* joint actions they will choose.

At this point we must clarify exactly what counts as an agent or coalition being able to perform some action (i.e., choose a valuation) which achieves their goal. Suppose for some $i \in A$ we have $\Phi_i = \{p\}$ and $\gamma_i = p \wedge \neg q$. Now, it might appear that i is able to achieve its goal in isolation, through the valuation $\{p\}$. However, *this is not the case*, since the achievement of γ_i depends upon the agent that controls q setting it to false. Thus, the utility obtained by agents within a coalition depends not just on their actions, but potentially on the actions of all agents in the game.

EXAMPLE 1. *Consider a game where we have two agents ($A = \{1, 2\}$) who can visit places B and S (game theorists may want to think of B as a Bach concert and S as a Stravinsky concert, although our example is* not *the same as the Bach and Stravinsky game that appears in the literature). Agent i going to B is represented by setting b_i to true, whereas his trip to S is represented by s_i. Hence we have $\Phi_i = \{b_i, s_i\}$. For agent 1 it is easier to*

go to S, whereas 2 lives close to B: $c(b_1) = c(s_2) = 2$ and $c(b_2) = c(s_1) = 1$. Note that $\mu = 6$. Regarding possible goals, we will look at 5 different agent types. Let i be an agent, and $j \neq i$:
DON'T CARE *(Don) has no constraints: $\gamma_i = \top$;*
FRIEND *(Fri) prefers to meet with the other: $\gamma_i = (b_i \wedge b_j) \vee (s_i \wedge s_j)$;*
FOE *(Foe) wants to go out without meeting the other agent: $\gamma_i = (b_i \wedge \neg b_j) \vee (s_i \wedge \neg s_j)$;*
UNREALISTIC *(Unr) has $\gamma_i = \bot$ as his goal;*
SOLIPSISTIC *(Sol) just wants to go out: $\gamma_i = (b_i \vee s_i)$.*

Based on these types $\{Don, Fri, Foe, Unr, Sol\}$ we can specify 25 types of games. For instance $G(Don, Fri)$ is the game in which agent 1 doesn't care about the outcome, but agent 2 wants to be a friend. Note that Don and Sol have a non-empty set of strategies for their goals. Let us say that agent i is happy given a valuation ξ if $\xi \models \gamma_i$, i.e., if i's goal is satisfied. We will, for this example, present valuations as $\mathsf{uvyz} \in \{0,1\}^4$, where u represents the value of b_1, v that of s_1, y is the value of b_2 and z that of s_2.

3. THE CORE

We say a valuation ξ_1 is *blocked by a coalition $C \subseteq A$ through a valuation ξ_2* iff:

1. ξ_2 is a feasible objection by coalition C:
$$\xi_2 = \xi_1 \mod C.$$
2. coalition C strictly prefers ξ_2 over ξ_1:
$$\text{for all } i \in C\text{: } \xi_2 \succ_i \xi_1.$$

Thus, if C blocks ξ_1 through ξ_2, then this means that C could do better than ξ_1 simply by flipping the value of some of the variables under their control. The *core* is the set of valuations that are not blocked by any coalition. Let $core(G)$ denote the core of G. First, we establish some general properties of the core of CBGs.

PROPOSITION 1. *Let $G = \langle A, \Phi, c, \gamma_1, \ldots, \gamma_n, \Phi_1, \ldots, \Phi_n \rangle$ be a game. Then:*

1. *$\xi \in core(G) \Rightarrow contrib(\xi) \subseteq ben(\xi)$*
2. *$\emptyset \in core(G) \Rightarrow (\emptyset \models \bigvee_{i \in A} \gamma_i$ or $core(G) = \{\emptyset\})$*
3. *$\xi \in core(G) \Rightarrow \xi$ is $contrib(\xi)$ minimal for $\gamma_{contrib(\xi)}$.*

Now, there are several obvious computational questions to ask with respect to $core(G)$. The first two of these are standard questions to ask of coalitional games in general:

> CORE MEMBERSHIP:
> *Given*: CBG G, valuation $\xi \subseteq \Phi$.
> *Question*: Is it the case that $\xi \in core(G)$?
>
> CORE NON-EMPTY:
> *Given*: CBG G.
> *Question*: Is it the case that $core(G) \neq \emptyset$?

EXAMPLE 1 (CONTINUED). *Let us first consider cases where both agents are of the same type. In the $G(Don, Don)$ game, the core is $\{0000\}$. This is intuitive: under this valuation, everybody is happy, and deviating from it would incur a cost for someone. In the $G(Fri, Fri)$ game, where both goals are $(b_1 \wedge b_2) \vee (s_1 \wedge s_2)$, the core is $\{0101, 1010\}$. Note that these are minimal valuations with the property that both agents are happy. We have $core(G(Foe, Foe)) = \{0110\}$: this is a valuation in which everybody is happy while maximising utility. And $core(G(Unr, Unr)) = \{0000\}$: since the agents' goals can neither be fulfilled, they better settle for incurring no cost. It is easy to see that $core(G(Sol, Sol)) = \{0110\}$. Moving on two mixed games (possibly different types) we have in fact that if one agent is a Don or an Unr type, he has no incentive to make any of his variables true. This is not good for a $Friend$ who needs cooperation from the other agents to satisfy his goals. We have, for any $Typ, Typ' \in \{Don, Unr\}$, that $core(G(Typ, Typ')) = core(G(Typ, Fri)) = \{0000\}$. However, Foe agents can benefit from Don and Unr agents, and Sol agents don't care: for all $Typ \in \{Don, Unr\}$, $core(G(Typ, Foe)) = \{0010\} = core(G(Typ, Sol))$. Note that the core can be empty, e.g., we have $core(G(Fri, Foe)) = \emptyset$ – for any valuation ξ, we can always find an agent who prefers a different valuation ξ'. For instance, note that $0000 \succ_2 0001 \succ_1 0101 \succ_2 0100 \succ_1 0000$, and every valuation is involved in such a chain with length > 1. We furthermore have $core(G(Fri, Sol)) = \{1010\}$ and, finally, $core(G(Foe, Sol)) = \{0110\}$*

THEOREM 1. CORE MEMBERSHIP *is co-*NP*-complete, even in games with a single agent, and even when the valuation to be checked is empty.*

PROOF. Membership of co-NP is clear from the statement of the problem. For hardness, we reduce SAT to the complement of the problem, i.e., the problem of determining whether a valuation is blocked. Let Ψ be the SAT instance, with Boolean variables $x_1, \ldots, x_k$. We create a game G_Ψ, as follows. We create a single agent, a_1, and let $\Phi = \{x_1, \ldots, x_k, d\}$, where d is a new Boolean variable, not occurring in Ψ. Then define $\gamma_{a_1} = \Psi \wedge d$, fix $c(v) = 1$ for all $v \in \Phi$, fix $\Phi_{a_1} = \Phi$, and fix $\xi = \emptyset$. We claim that $\xi \notin core(G_\Psi)$ iff Ψ is satisfiable. The proof follows immediately from construction. $\square$

We note that, if we use the cost function $c(v) = 0$ for $v \in \Phi \setminus \{d\}$ and $c(d) = 1$, then every satisfying instantiation of Ψ maps to a distinct valuation in $core(G_\Psi)$. We thus derive,

COROLLARY 1. *Given a* CBG*, G, with $c(v) \in \{0, 1\}$ for each $v \in \Phi$, computing $|core(G)|$ is #*P*–hard.*

Theorem 1 considers the special case of the *empty* evaluation. Informally, one could view this case as asking whether taking *no action at all* is a justifiable collective strategy. Suppose one considers a similar question, namely, if every agent executes every action under its control, is it possible for any coalition to improve on the resulting outcome, i.e., is the valuation Φ in the core? Even in this case, we have:

COROLLARY 2. CORE MEMBERSHIP *is co-*NP*-complete, even in games with a single agent and where the valuation to be checked is Φ.*

PROOF. Use a similar reduction from SAT to the complementary problem, but with $\gamma_{a_1} = \Psi \wedge (\neg d) \vee (\wedge_{v \in \Phi} v)$. $\square$

There are, however, cases for which the core membership problem of Corollary 2 can be be decided efficiently. A goal, γ, is said to be Φ-*positive* if the $\mathcal{L}$-formula over Φ that defines γ is constructed using only operators from $\{\wedge, \vee\}$. An easily seen property of Φ-positive goals, γ, is: if $\zeta \in [\![\gamma]\!]$ then $\xi \in [\![\gamma]\!]$, for every $\xi \supset \zeta$.

THEOREM 2. *Let G be a* CBG *in which each γ_i is Φ-positive. Deciding if $\Phi \in core(G)$ can be carried out in polynomial time.*

PROOF. Given G as in the theorem statement, first observe that $\Phi \notin \llbracket \gamma_i \rrbracket$ if and only if $\gamma_i \Leftrightarrow \bot$: if $c(\Phi_i) > 0$ then $\Phi \notin core(G)$. So, without loss of generality, in testing $\Phi \in core(G)$, we may focus attention on those $a_i \in A$ for whom $\Phi \in \llbracket \gamma_i \rrbracket$. Suppose C blocks Φ through a valuation ζ. It is easy to see that for each $a_i \in C$ there is some valuation ζ_i that blocks Φ: the only way in which $\zeta \succ_i \Phi$ for each $a_i \in C$ is for a_i not to perform some action under its control while retaining the property of its goal being satisfied.

Using the observations above we can test $\Phi \in core(G)$ as follows: first check if there is any a_i for which $\Phi \notin \llbracket \gamma_i \rrbracket$ and $c(\Phi_i) > 0$. If this is the case then $\Phi \notin core(G)$ as it is blocked by $\{a_i\}$ through the valuation $\Phi \setminus \Phi_i$. Otherwise, for each $x \in \Phi$ with $c(x) > 0$ check whether $\Phi \setminus \{x\} \in \llbracket \gamma_i \rrbracket$ where a_i is the agent controlling x. Again, if there is such an x then Φ is blocked by $\{a_i\}$ through the valuation $\Phi \setminus \{x\}$. If no suitable x is identified then $\Phi \in core(G)$. □

THEOREM 3. CORE NON-EMPTY *is* Σ_2^p*-complete.*

PROOF. Membership is straightforward from the problem definition. For hardness, we reduce the problem of determining whether $\text{QBF}_{2,\exists}$ formulae are true [12]. An instance of $\text{QBF}_{2,\exists}$ is given by a quantified Boolean formula with the following structure:

$$\exists \bar{x} \; \forall \bar{y} \; \chi(\bar{x}, \bar{y}) \tag{1}$$

in which $\bar{x}$ and $\bar{y}$ are (disjoint) sets of Boolean variables, and $\chi(\bar{x}, \bar{y})$ is a propositional logic formula (the *matrix*) over these variables. Such a formula is true if there exists an assignment ξ_1 to $\bar{x}$ such that for all assignments ξ_2 to $\bar{y}$, we have $\xi_1 \cup \xi_2 \models \chi(\bar{x}, \bar{y})$. An example of a $\text{QBF}_{2,\exists}$ formula is:

$$\exists x \forall y [(x \vee y) \wedge (x \vee \neg y)] \tag{2}$$

This formula is in fact clearly true, as witnessed for example by the existential variable assignment $\{x\}$. Let $\bar{x} = \{x_1, \ldots, x_g\}$ be the universally quantified variables in the input formula, let $\bar{y} = \{y_1, \ldots, y_h\}$ be the existentially quantified variables, and let $\chi(\bar{x}, \bar{y})$ be the matrix.

The construction is in 2 parts. The first part directly corresponds to the input formula (1), as follows:

- create $g + h$ agents: $A = \{1, \ldots, g + h\}$;
- fix $\Phi = \bar{x} \cup \bar{y} \cup \{d\}$, where d is a new variable, not appearing in $\bar{x} \cup \bar{y}$;
- fix $\Phi_1 = \{x_1, d\}$ and for all $2 \leq i \leq g$, fix $\Phi_i = \{x_i\}$;
- for all $1 \leq i \leq g$ fix $\gamma_i = \chi(\bar{x}, \bar{y}) \wedge d$
- for all $g + 1 \leq i \leq g + h$, fix $\gamma_i = (\neg \chi(\bar{x}, \bar{y})) \wedge d$ and $\Phi_i = \{y_{i-g}\}$; and finally,
- fix $c(v) = 1$ for each $v \in \Phi$.

Let $A_\exists$ be the agents corresponding to existentially quantified variables, and let $A_\forall$ be the agents corresponding to universally quantified variables. Thus, $A_\exists$ want to make the matrix true (as well as d), while $A_\forall$ want to make it false (while making d true). In fact, $A_\forall$ will never have a joint action to make their goal true, as they do not control d.

The second part of the construction ensures that the core does not contain $\emptyset$ in the event that the input formula (1) is false. We create 3 additional agents, $\delta_1, \delta_2, \delta_3$, and 6 additional variables $\zeta_1, \ldots, \zeta_6$, with controlled variables as follows: $\Phi_{\delta_1} = \{\zeta_1, \zeta_2\}$, $\Phi_{\delta_2} = \{\zeta_3, \zeta_4\}$, and $\Phi_{\delta_3} = \{\zeta_5, \zeta_6\}$. Goals for the additional agents are defined in two parts, as follows. First, we define *auxilliary* goal formulae, ρ_i, as follows: $\rho_{\delta_1} = (\zeta_3 \vee \zeta_6)$, $\rho_{\delta_2} = (\zeta_2 \vee \zeta_5)$, and $\rho_{\delta_3} = (\zeta_1 \vee \zeta_4)$. We then define the goal formulae as follows: $\gamma_{\delta_1} = \chi(\bar{x}, \bar{y}) \vee (\rho_{\delta_1} \wedge \neg(\rho_{\delta_2} \wedge \rho_{\delta_3}))$, $\gamma_{\delta_2} = \chi(\bar{x}, \bar{y}) \vee (\rho_{\delta_2} \wedge \neg(\rho_{\delta_1} \wedge \rho_{\delta_3}))$, and $\gamma_{\delta_3} = \chi(\bar{x}, \bar{y}) \vee (\rho_{\delta_3} \wedge \neg(\rho_{\delta_1} \wedge \rho_{\delta_2}))$. Finally, the cost function for the additional variables is defined as follows: $c(\zeta_1) = 2$, $c(\zeta_2) = 1$, $c(\zeta_3) = 2$, $c(\zeta_4) = 1$, $c(\zeta_5) = 2$, and $c(\zeta_6) = 1$.

Let G be the game thus constructed. We claim that (1) is true iff $core(G) \neq \emptyset$. The key difficulty in the proof is in showing that $\emptyset \notin core(G)$ if (1) is false; the second part of the construction above handles this case. □

The next questions to ask, however, are specifically tailored to CBGs. We are given a propositional formula $\varphi \in \mathcal{L}$, and asked whether, *no matter which outcome in the core were chosen*, this outcome would satisfy φ. More formally, the decision problem is:

> UNIVERSAL CORE PROPERTY:
> *Given*: CBG G, formula $\varphi \in \mathcal{L}$.
> *Question*: Is it the case that $core(G) \subseteq \llbracket \varphi \rrbracket$?

THEOREM 4. UNIVERSAL CORE PROPERTY *is* Π_2^p*-complete.*

PROOF. We deal with the complement problem, i.e., the problem of deciding whether $\exists \xi \in core(G) : \xi \not\models \varphi$. Membership of Σ_2^p is clear from the problem statement. For hardness, reduce CORE NON-EMPTY: in the construction, we leave the game unchanged, and simply define the property to be checked as $\varphi = \bot$. Correctness of the reduction is immediate. Since the complement problem is Σ_2^p-complete, UNIVERSAL CORE PROPERTY is Π_2^p-complete. □

The obvious EXISTENTIAL CORE PROPERTY problem asks whether $\exists \xi \in core(G) : \xi \models \varphi$. Using the same proof idea as Theorem 4, but defining $\varphi = \top$, we immediately get:

COROLLARY 3. EXISTENTIAL CORE PROPERTY *is* Σ_2^p*-complete.*

We can consider CORE CONTAINMENT, the converse direction to UNIVERSAL CORE PROPERTY.

> CORE CONTAINMENT:
> *Given*: CBG G, formula $\varphi \in \mathcal{L}$.
> *Question*: Is it the case that $\llbracket \varphi \rrbracket \subseteq core(G)$?

Perhaps surprisingly, this problem turns out to be "easier" (under standard complexity theoretic assumptions) than the closely related UNIVERSAL CORE PROPERTY problem.

THEOREM 5. CORE CONTAINMENT *is co-*NP*-complete even if instances are restricted to* $\langle G, \varphi \rangle$ *with* φ *a* Φ*-positive formula.*

PROOF. Membership of co-NP is immediate from the problem statement. For hardness, we use the result of Corollary 2 that deciding if $\Phi \in core(G)$ is co-NP–complete. Given an instance, G, of this problem we leave G unchanged and define $\varphi = \bigwedge_{v \in \Phi} v$. Correctness of the reduction is immediate by construction. □

Finally, we might also want to consider whether a property $\varphi \in \mathcal{L}$ *characterises* the core, in the following sense:

> CORE CHARACTERISATION:
> *Given*: CBG G, formula $\varphi \in \mathcal{L}$.
> *Question*: Is it the case that $\forall \xi \subseteq \Phi$, we have $(\xi \in core(G))$ iff $(\xi \models \varphi)$?

The following is immediate from the results above.

COROLLARY 4. CORE CHARACTERISATION *is* Π_2^p*-complete.*

EXAMPLE 1 (CONTINUED). *Consider the following property, which we refer to as* $coordination$: $(b_1 \leftrightarrow b_2) \wedge (s_1 \leftrightarrow s_2)$. *Of the games mentioned earlier, this property is universal for any game in which one of the player is of type* Don *or* Unr*, and also for* $G(Fri, Foe)$*, since it has an empty core.* coordination *is furthermore universal for* $G(Fri, Fri)$ *and* $G(Fri, Sol)$*. Core containment holds for no game we have discussed, in fact not even for the game* G *in which both agents' goal would be* coordination*: note that* $[\![coordination]\!] = \{0000, 0101, 1010, 1111\}$*, and that the* $core(G) = \{0000\}$.

Next, consider the problem of *comparing* CBGs (cf. discussion in [17, Section 4]). For CBGs G_1 and G_2 with identical sets of agents and actions, one natural definition of equivalence is that $core(G_1) = core(G_2)$. It is not difficult to see that given $\langle G_1, G_2 \rangle$, deciding if $core(G_1) = core(G_2)$ in this way is co-NP-complete: for $A = \{a\}$, define $\gamma_a^{(1)} = \top$ and $\gamma_a^{(2)} = \psi$ for input formula ψ that we wish to check for tautology status. Using $c_1(v) = c_2(v) = 0$ for each $v \in \Phi$, one then has $core(G_1) = core(G_2)$ iff ψ is a tautology.

Finally, we show that any game with a natural number cost function can be "simulated" by one which uses a cost function using only values from $\{0, 1\}$ with only a modest increase in the "size" of the game. Let $G = \langle A, \Phi, c, \gamma_1, \ldots, \gamma_n, \Phi_1, \ldots, \Phi_n \rangle$, with $c : \Phi \to \mathbb{N}$. We define the *size of* G as,

$$size(G) = \sum_{i=1}^{n} |\gamma_i| \ + \ (\lceil \log_2 r + 1 \rceil)|\Phi|$$

where $r = \max_{v \in \Phi} \ c(v)$.

THEOREM 6. *Let* G *be a* CBG *with an integer valued cost function whose maximum value is* r_c*. There is a* CBG,

$$H = \langle A, \Phi, c', \gamma'_1, \ldots, \gamma'_n, \Phi'_1, \ldots, \Phi'_n \rangle$$

with the following properties

1. *There are polynomial time computable mappings,*

$$\tau_1 \ : \ 2^{\Phi} \to 2^{\Phi'}$$
$$\tau_2 \ : \ 2^{\Phi'} \to 2^{\Phi}$$

   for which

$$\xi \in core(G) \Leftrightarrow \tau_1(\xi) \in core(H)$$
$$\xi \in core(H) \Leftrightarrow \tau_2(\xi) \in core(G).$$

2. $size(H) = O(r_c \ size(G))$.

3. $c'(v) \in \{0, 1\}$ *for all* $v \in \Phi'$.

PROOF. (Sketch) Given G as in the theorem statement, if $v \in \Phi$ has $c(v) = 0$ then add v to Φ and set $c'(v) = 0$. For each $v \in \Phi$ with $c(v) = k > 0$, replace v by k actions $\{v_1, \ldots, v_k\}$ in Φ' with $c'(v_j) = 1$. To form γ'_i, substitute the conjunction $(\wedge_{j=1}^{k} v_j)$ for each occurrence of v in γ_i (when $c(v) > 0$). We then define $\tau_1(\xi)$ by replacing each $v \in \xi$ with the corresponding *set* from Φ'; similarly $\tau_2(\xi)$ replaces $w \in \xi$ by the action in Φ from which w arose. Correctness of the translations follow from construction; we omit the remaining technical details for space reasons. □

4. STABLE SETS

While the core is the most studied solution concept in cooperative games, it is not the oldest: this honour goes to *stable sets* [11, pp.278–281], which were originally introduced by von Neumann and Morgenstern [10]. Crudely, a stable set is a set of game outcomes, which have the property that they are both *internally* and *externally* stable: internal stability means that no member of the set is preferred over another member of the set, while external stability means that for every element outside the set, there is some element within the set that is preferred over it.

More formally, let G be a CBG, and let $X \subseteq 2^{\Phi}$ be a set of valuations. Then X is a stable set of G iff it satisfies the following properties (cf. [11, p.279]):

Internal stability: If $\xi \in X$ then for no $C \subseteq A$ does there exist a $\xi' \in X$ such that C objects to ξ through ξ'.

External stability: If $\xi \in (2^{\Phi} \setminus X)$ then there exists a $\xi' \in X$ and $C \subseteq A$ such that C objects to ξ through ξ'.

Now, in conventional coalitional games, it is known that there may be multiple stable sets, and each stable set can contain many outcomes; moreover, there may be no stable sets [11, p.279].

EXAMPLE 1 (CONTINUED). *In the running example we have seen so far, for every* G, $core(G)$ *is a stable set. To see that for instance* $X = \{0101, 1010\}$ *is a stable set in* $G(Fri, Fri)$*, we first note that* X *is internally stable: only the grand coalition* $B = \{1, 2\}$ *has the property that* $0101 = 1010 \mod B$ *(no agent can force a change from one valuation to the other) and for this coalition, the valuations are incomparable. For external stability, take a valuation* $\xi \notin X$*. Then* (i) *either it makes (at least) one more atom true (the agent whose atom it is will object: he makes an unnecessary cost) or* (ii) *it makes at least one more atom false (both agents will object, their goal is not satisfied anymore).*

In general, stable sets don't have to coincide with the core, though. Consider our final example game, based on our running example. Suppose $\gamma_1 = (b_1 \vee s_1) \wedge (b_2 \to b_1)$ *and* $\gamma_2 = \neg b_1 \to b_2$*. (Agent 1 likes to go out, and if agent 2 goes to Bach, 1 is willing to make the additional cost and go there as well; agent 2 likes Bach, but goes there if 1 will not). Call this game* H*. Then we claim:*

1. $core(H) = \emptyset$*: Every valuation that satisfies* $b_1 \wedge s_1$ *is blocked by agent 1: if such a valuation satisfies* b_2 *as well, agent 1 better choose* $b_1 \wedge \neg s_1$ *(satisfying his goal and minimising costs), otherwise he chooses* $\neg b_1 \wedge s_1$ *(idem). Similarly, valuations satisfying* $\neg b_1 \wedge \neg s_1$ *are blocked by agent 1: if agent 2 chooses* b_2*, agent 1 better choose* $b_1 \wedge \neg s_1$*, otherwise agent 1 will prefer* $\neg b_1 \wedge s_1$*. Among the valuations satisfying* $\neg b_1 \wedge s_1$*, agent 2 strictly prefers* 0110*, hence block all the others. However,* 0110 *on its turn is blocked by agent 1: he prefers* 1010*. Finally, look at valuations satisfying* $b_1 \wedge \neg s_1$*. Among them, agent 2 strictly prefers* 1000 *and hence blocks all the others. But then again, agent 1 strictly prefers* 0100 *over* 1000*, so that no valuations stays unblocked by some coalition.*

2. *The following are two stable sets:* $X_1 = \{0110, 1000\}$ *and* $X_2 = \{0100, 1010\}$*. It is easy to see that both sets are internally stable. For external stability, note that we have* $1000 \succ_1 0100 \succ_2 0110 \succ_1 1010 \succ_2 1000$*, so that indeed, every member from* X_1 *is blocked by some member of* X_2*, and vice versa. It remains to be demonstrated that every valuation outside* X_1 *and* X_2 *is blocked by both a valuation from* X_1 *and one from* X_2*. We do this for* X_1*. Note that*

$u_1(1000) = 4$ and $u_2(1000) = 6$. All other valuations are strictly worse for both agents, except for the following three cases: (i) 0100, for which $\mu_1(0100) = 5$, but we have already argued that $0100 \in X_2$ is blocked by some valuation in X_1; (ii) 0101, with $\mu_1(0101) = 5$: however we have that agent 2 objects through 0110 against 0101; (iii) 1100, which gives agent 2 a utility of 6. However, agent 1 objects through 1000 against 1100.

From a computational point of view, the most obvious issue arising is that of *succinctly representing* stable sets for CBGs. That is, since a stable set is a subset of 2^Φ, it could be that such a set is exponentially large in the size of Φ, and thus representing a stable set directly (by explicitly enumerating its contents) is not feasible. (If we make the unrealistic assumption of an explicit enumeration, then checking both internal and external stability are easily seen to be solvable in polynomial time in the $|X|$, i.e., exponential in $|\Phi|$.) So, instead of representing X explicitly, we assume X is represented as a formula $\varphi_X \in \mathcal{L}$, with the assignments that satisfy φ_X corresponding to the members of X. This gives us the following problem:

INTERNAL STABILITY:
Given: CBG G, $\varphi_X \in \mathcal{L}$.
Question: Is $[\![\varphi_X]\!]$ internally stable?

The EXTERNAL STABILITY and STABLE SET problems are then defined in the obvious way.

THEOREM 7. INTERNAL STABILITY *is co-*NP*-complete.*

PROOF. Consider the complement problem. Membership of NP is obvious. For hardness, we reduce SAT. Given a SAT instance Ψ, over variables $x_1, \ldots, x_k$, we create a game G_Ψ with one agent, $A = \{1\}$, $\Phi = \{x_1, \ldots, x_k, d\}$, define $c(x_i) = 0$ for all $1 \leq i \leq k$, and define $c(d) = k + 1$. Finally, define

$$\gamma_1 = \varphi_X = \Psi \vee (d \wedge \bigwedge_{i=1}^{k} \neg x_i).$$

If Ψ is satisfiable, then $[\![\varphi_X]\!]$ will contain the valuation $\{d\}$ with $c_i(\{d\}) = k + 1$, together with all valuations ξ that satisfy Ψ, and for each such ξ, $c_i(\xi) \leq k$, in which case $[\![\varphi_X]\!]$ will not be internally stable. If Ψ is unsatisfiable, then the only valuation in $[\![\varphi_X]\!]$ will be $\{d\}$, which will thus be internally stable. Since the complement is NP-complete, INTERNAL STABILITY is co-NP-complete. $\square$

THEOREM 8. EXTERNAL STABILITY *is Π_2^p-complete.*

PROOF. Membership is immediate from from the problem definition. To show EXTERNAL STABILITY is Π_2^p-hard, we use a reduction from CORE EMPTINESS, i.e. the complementary problem to that shown Σ_2^p-hard in Theorem 3. Given an instance G of CORE EMPTINESS, form the instance $\langle H, \psi \rangle$ of EXTERNAL STABILITY in which $\Phi_H = \Phi_G \cup \{z\}$ with z a new action. Fix $c(z) = 0$ and, without loss of generality, let z be controlled by a_1 in H. No other changes to G are made. To complete the instance, we set $\psi = \neg z$. We now argue that G has an empty core if and only if $\langle H, \neg z \rangle$ is externally stable. Suppose $core(G) = \emptyset$. The set of valuations that are *not* in $[\![\psi]\!]$ take the form $\{z\} \cup \xi$ with $\xi \subseteq \Phi$. Since every $\xi \subseteq \Phi$ is blocked by some C via a valuation $\xi' \subseteq \Phi$, we deduce that C objects to $\{z\} \cup \xi$ via the evaluation $\xi' \in [\![\psi]\!]$, i.e., if $core(G)$ is empty then ψ is externally stable w.r.t. H. On the other hand, suppose that $core(G) \neq \emptyset$. Consider any $\xi \in core(G)$ and the valuation $\{z\} \cup \xi \notin [\![\neg z]\!]$. Suppose that some coalition, C, objects to $\{z\} \cup \xi$ via a valuation ξ'. Since $c(z) = 0$ and does not influence whether any goal is realised, it follows that this same coalition has a feasible objection to ξ via $\xi' \setminus \{z\}$. This contradicts $\xi \in core(G)$. It follows that $\{z\} \cup \xi$ cannot be blocked by any $\langle C, \xi' \rangle$, i.e., ψ is not externally stable w.r.t. H. $\square$

The following is now straightforward.

COROLLARY 5. STABLE SET *is Π_2^p-complete.*

5. A BARGAINING PROTOCOL

In the preceding sections, we investigated two types of solutions for Boolean games, mainly concerned with checking whether a particular valuation is stable against defections. In this section, we address the issue of the method by which a group of agents can agree upon a valuation. Our approach is to present a *negotiation protocol* for cooperation in Boolean games. Although negotiation protocols have long been studied in multi-agent systems research (see e.g., [13, 9]), our approach is novel in respect to three aspects. First, most research on multi-agent systems has focused on *bilateral* negotiations, that is, negotiations between two agents: we present a protocol for n agent negotiations. Second, almost all research on negotiations in multi-agent systems has assumed transferable utility: we assume *non-transferable utility*. In the game theory community, much less research has been devoted to bargaining in the NTU case than the TU case, and even less research in the multi-agent systems domain has considered NTU bargaining. Third, most multi-agent research has assumed that the utility obtained by agents within a coalition does not depend on the actions of all agents in the game. Also in game theory, most research on coalition formation has been on bargaining without externalities: we study situations where the actions of agents in one coalition may influence the utility of agents in other coalitions.

We present a protocol that will be able to identify strategies in equilibrium. The protocol is described in Algorithm 1. The intuition behind this protocol is that it balances the power given to the proposers and the responders, as follows. First, the responders can reject the offer, in which case the proposer gains nothing from negotiation. Thus, the proposer must present a "beneficial" offer – if it proposes a poor offer, the others can reject it, forcing the proposer to leave the negotiations, and forgo the benefits of cooperation. However, the proposer has the power to choose *which* offer to make: between all of the beneficial offers that it could make, it can choose its most preferred one. Note that fairness is obtained by applying a randomized order.

Overall, the protocol directs the agents to a Pareto optimal solution, by allowing them to improve upon previous proposed valuations, given that the other agents will not suffer from such improvement. Another advantage of the protocol is that, once the order of the agents is determined, the protocol is deterministic. This makes it easier to compute the equilibrium strategies.

In the following analysis we will use the concept of subgame-perfect equilibrium which is the appropriate stability concept for games with several steps. First, a *bargaining strategy* is a function used by an agent to decide what action to take during negotiation (i.e., what offer to make, if it is this agent's turn to make an offer, whether to reject a proposal or not, and so on; we will not give a formal definition of bargaining strategies – see, e.g., [9, p.23]). A *strategy profile* is a sequence of such strategies, one for each agent. A strategy profile $(f_1, f_2, \ldots, f_n)$ is a *Nash equilibrium* if each agent i does not have a different strategy yielding an outcome that it prefers to that generated when it chooses f_i, given that

Algorithm 1 Negotiation Protocol for Boolean games

Input: $G = \langle A, \Phi, c, \gamma_1, \dots, \gamma_n, \Phi_1, \dots, \Phi_n \rangle$,

Randomly choose an ordering $\langle a_1, \dots, a_n \rangle$ of A.
{Negotiation stage begins}
$t := 0$; $\Phi^1 = \Phi$
repeat
  $accept := \mathbf{true}$
  $t := t + 1$;
  a_t proposes a valuation, $\xi^{t,t} \subseteq \Phi^t$
  $i := t$
  repeat
    $i := i + 1$
    if a_i rejects $\xi^{t,i-1}$ **then**
      $accept := \mathbf{false}$
      a_t opts out and plays the game
        $G^t = \langle \{a_t\}, \Phi_{a_t}, c, \gamma_{a_t}, \Phi_{a_t} \rangle$.
      $\Phi^{t+1} = \Phi^t \setminus \Phi_{a_t}$
      {Note: a_t only gains the utility achieved in the game G^t regardless of what valuation is agreed by the remaining agents.}
    else
      {a_i makes a counteroffer}
      a_i proposes $\xi^{t,i} \subseteq \Phi^t$, a proposal (possibly the same as the one put forward by its immediate predecessor in the ordering) that must satisfy $\bigwedge_{j=t}^{i-1} [\xi^{t,i} \succeq_j \xi^{t,i-1}]$
      {i.e. The proposal made by a_i cannot leave any agent a_j, preceding a_i in the ordering, *strictly* preferring its earlier proposal $\xi^{t,i-1}$ in round t.}
    end if
  until (**not** $accept$) **or** $(i = n)$
until $(t = n)$ **or** $(accept)$
$\xi^{t,n}$ is the agreed valuation.

the other players follow their profile strategies [11, p.14]. Finally, a strategy profile is a *subgame perfect equilibrium* if the strategy profile induced in every subgame is Nash equilibrium of that subgame [11, p.97]. This means that at any step of the negotiation process, no matter what the history is, no agent is motivated to deviate and use another strategy other than that defined in the strategy profile. Once the order $a_1, \dots, a_n$ of the agents is fixed, we can compute the strategies that are in perfect equilibrium using backward induction [11, p.99].

We denote by $u_i(\{i\})$ the value $\max_{\xi \subset \Phi_i} u_{a_i}(\xi)$. We denote $\bar{\xi}_i = \arg\max_{\xi \subset \Phi_i} u_{a_i}(\xi)$ (there may be several such valuations; we choose one arbitrarily). Other notations we use in the analysis are as follows: For $1 \le t \le n$, $u_i(t,j)$ denotes the utility that agent a_i obtains from a proposal made by agent a_j at round t, i.e., $u_i(t,j) = u_i(\xi^{t,j})$. We denote by $u_i(t)$ the utility that agent a_i expects to obtain at iteration t. If it is expected that an agreement will be reached in round t then $u_j(t) = u_j(t,n)$. If a_t is expected to opt out in round t then $u_t(t) = u_t(\{t\})$ and for $t < j \le n$ $u_j(t) = u_j(t+1)$.

Suppose, the negotiation has reached round t, Φ^t has been defined, and it is agent a_i's turn, $i = t$, to make an offer $\xi^{t,t}$ or its turn to make a counterproposal $\xi^{t,i}$, (where $t < i \le n$). Agent a_i is seeking to maximize $u_i(t,i)$, but does not have complete freedom (within the protocol defined by Algorithm 1) simply to propose a valuation which is solely in its own interest: to do so, when $i = t$, might lead to agents a_j rejecting its proposal outright $(j > i)$. Hence, a_i must propose valuations that a_j prefers to any that could be offered in future negotiations. Furthermore, the proposed valuation should be better for a_i than opting out or the utility obtained when acting in isolation. When a_i, $i > t$, makes an offer, Algorithm 1 does not allow $\xi^{t,i}$ to reduce the utility of the agents that have accepted $\xi^{t,j}$, $j < i$.[1] In addition, similarly to a_t it must make sure that the agents that respond throughout the remainder of the round a_j, $i < j \le n$ obtain a higher utility than their future one.

Formally, denote by $\Xi^{t,i}$ the set of valuations that agent a_i should consider. Any $\xi \in \Xi^{t,i}$ must satisfy the following constraints:

$$u_j(\xi) \ge u_j(t+1) \qquad i < j \le n \tag{3}$$

$$u_i(\xi) \ge \begin{cases} u_i(\{i\}) & i = t \\ u_i(t+1) & \text{otherwise} \end{cases} \tag{4}$$

$$u_j(\xi) \ge u_j(i-1,t) \qquad t \le j < i \tag{5}$$

We denote by $Max(t,i)$ the proposal obtained by a_i via the following rule: If $\Xi^{t,i} = \emptyset$ then a_i, $t < i \le n$ should reject the offer; otherwise, if $\Xi^{t,i} \ne \emptyset$, then a_i should propose the member of $\Xi^{t,i}$ that maximizes its utility.

$$Max(t,i) = \begin{cases} \text{argmax}_{\xi \in \Xi^{t,i}} u_{a_i}(\xi) & \Xi^{t,i} \ne \emptyset \\ \bar{\xi}_t & \text{otherwise} \end{cases} \tag{6}$$

In order to identify the strategies that are in perfect equilibrium the agents need to compute $\Xi^{i,t}$. Assume that the current round is t and it is agent a_i's turn to propose a valuation. The value of $u_j(i-1,t)$ that appears in constraint (5) for each agent j preceding i according to the predefined order, i.e. those for which $t \le j \le i-1$, can be calculated easily from the proposal agent a_{i-1} has already made.

However, the computation of the value $u_j(t+1)$, for $i \le j \le n$, that appears in constraints 3 and 4, requires computing the values of $u_j(t')\ \forall t' > t$ and $\forall t' \le j \le n$, i.e. solutions of $Max(t,i)$ are specified in terms of the solutions of $Max(t',j)$ for every $t' > t$ and $t' \le j \le n$.

We resolve this difficulty as follows: to compute the required values the agent uses a backward induction starting from computing $u_n(n)$ by finding a solution for $Max(n,n)$ which will lead to $u_n(\{n\})$. Then the agent computes $u_{n-1}(n-1)$ and $u_n(n-1)$ by finding solutions for $Max(n-1,n-1)$ and $Max(n-1,n)$. In general, given a round $t' > t$ the agent computes $u_j(t',i)$ by finding a solution for $Max(t',i)$ according to the agents predefined order (i.e. it begins solving $u_j(i,t')$ for $i = t'$ till $i = n$). Note that when the agent calculates $u_j(t',i)$, the values of $u_j(t',k)$ have already been calculated for $t' \le k \le j-1$ and the values of $u_k(t'+1)$ have already been computed for $t'+1 \le k \le n$.

These processes are formally presented in the following theorem.

THEOREM 9. *Given an order of the agents* $(a_1, \dots, a_n)$*, the following strategies are in subgame-perfect equilibrium:*

For any round $1 \le t < n$,

- *Agent* a_t *offers the valuation obtained by solving* $Max(t,t)$.
- *Agent* a_i, $t < i \le n$ *on its turn in round* t *will:*

  Given the valuation that has already been proposed by a_{i-1}, *compute* $u_j(t, i-1)$, $t \le j < i$.

[1] i.e. "accepted" in the sense that, since a_i is still a participant, no agent a_j with $t < j < i$ has explicitly rejected $\xi^{t,l}$, $t \le l < i$.

Then, it will compute $\Xi^{t,i}$. If $\Xi^{t,i} = \emptyset$ then a_i will "reject" the offer. Otherwise, a solution for $Max(t, j)$ will be computed as described above and a_i will offer the valuation obtained.

PROOF. (Sketch) By backward induction on the negotiation step (t), and for each step by backward induction according to the agent's order. □

Two important questions must be considered with respect to the results of the negotiations: (i) When will the negotiations end? and (ii) Will the results be Pareto Optimal? In general, the negotiations may last for several rounds, and may even end only at the n'th round. For example, consider a game where there are three agents: 1, 2, and 3. Suppose agent 1 and agent 2 control variables p and q, respectively. Also suppose that agent 1's goal is q, agent 2's goal is p and agent 3's goal is $\neg p \wedge \neg q$. If the order of the agents is $a_1 = 1$, $a_2 = 2$ and $a_3 = 3$, then it is easy to see that the negotiations will end after 3 rounds and each agent will work in isolation, even though both 1 and 2 could increase their utility by forming a coalition and agreeing upon the valuation $\{p, q\}$. However, if the order is $a_1 = 3$, $a_2 = 2$ and $a_3 = 1$ then agent 3 will leave the negotiations and in the second round agents 1 and 2 will agree upon the valuation $\{p, q\}$. Note that the set of goals in this example is inconsistent. However, even if the set of goals is consistent the agents may not be able to agree upon a valuation. For example, suppose agent 1 controls variable p and q, $c(p) = 20$ and $c(q) = 1$, its goal is $p \vee q$ and agent 2's goal is $\neg q$.

However, if we consider CBGs in which each γ_i is Φ-positive then the negotiations will end during the first round with a Pareto optimal valuation. For example, suppose agent 1 controls variable p, $c(p) = 20$ and its goal is $p \vee q$, agent 2 controls the variable q, $c(q) = 20$ and its goal is $q \vee w$ and agent 3 controls w, $c(w) = 20$ and its goal is $p \vee w$. Regardless of the order of the agents in the negotiations they will end during the first round. For both orders $a_1 = 1$, $a_2 = 2$ and $a_3 = 3$ and $a_1 = 3$, $a_2 = 2$ and $a_3 = 1$ the agreed upon valuation in the first round will be $\{p, w\}$. For the order $a_1 = 2$, $a_2 = 1$ and $a_3 = 3$ the agreed upon valuation in the first round will be $\{q, w\}$. Finally, for the order $a_1 = 1$, $a_2 = 3$ and $a_3 = 2$ the agreed upon valuation in the first round will be $\{p, q\}$.

THEOREM 10. *Let G be a* CBG *in which each γ_i is Φ-positive. If the agents follow our perfect equilibrium defined strategies then the negotiations will end during the first round and any valuation resulting from the negotiations using our perfect equilibrium defined strategies will be Pareto optimal.*

PROOF. (Sketch) By induction on the number of agents n. □

6. CONCLUSIONS

Cooperative games present many challenges from the point of view of multi-agent systems research: how to represent them is one key challenge, and how to compute solution concepts for them is another. In this paper, we have presented a novel model for cooperative games, which is both compact and expressive. It provides a very natural framework through which to understand collective action in systems where agents have goals to achieve, and the actions available to agents have some cost. There are very many interesting questions for future work. The most obvious is to what extent we can identify sub-classes of CBGs for which the computation of solution concepts is tractable; another is to investigate in more detail the interplay between solution concepts, the logical structure of goal formulae, and the propositions controlled by participant agents. It is also interesting to characterise the class of coalitional games our cooperative Boolean games exactly correspond to. A natural extension would be to add *overall constraints* to the system (expressing, for instance, that an agent will not go *both* to Bach and Stravinsky), and also to incur costs for setting a proposition to *false*. The latter two issues could be investigated in tandem, one way to administer costs for a proposition to become false would be to associate it with another proposition with opposite truth-value.

7. REFERENCES

[1] E. Bonzon. *Modélisation des interactions entre agents rationnels : les jeux booléens*. PhD thesis, Université Paul Sabatier, Toulouse, 2007.

[2] E. Bonzon, M. Lagasquie, J. Lang, and B. Zanuttini. Boolean games revisited. In *Proceedings of the Seventeenth European Conference on Artificial Intelligence (ECAI-2006)*, Riva del Garda, Italy, 2006.

[3] V. Conitzer and T. Sandholm. Complexity of constructing solutions in the core based on synergies among coalitions. *Artificial Intelligence*, 170:607–619, 2006.

[4] X. Deng and C. H. Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of Operations Research*, 19(2):257–266, 1994.

[5] P. Harrenstein. *Logic in Conflict*. PhD thesis, Utrecht University, 2004.

[6] P. Harrenstein, W. van der Hoek, J.-J. Meyer, and C. Witteveen. Boolean games. In J. van Benthem, editor, *Proceeding of the Eighth Conference on Theoretical Aspects of Rationality and Knowledge (TARK VIII)*, pages 287–298, Siena, Italy, 2001.

[7] W. Hoek and M. Wooldridge. On the logic of cooperation and propositional control. *Artificial Intelligence*, 164(1-2):81–119, May 2005.

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

[9] S. Kraus. *Strategic Negotiation in Multiagent Environments*. The MIT Press: Cambridge, MA, 2001.

[10] J. Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press: Princeton, NJ, 1944.

[11] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. The MIT Press: Cambridge, MA, 1994.

[12] C. H. Papadimitriou. *Computational Complexity*. Addison-Wesley: Reading, MA, 1994.

[13] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. The MIT Press: Cambridge, MA, 1994.

[14] T. Sandholm, K. Larson, M. Andersson, O. Shehory, and F. Tohmé. Coalition structure generation with worst case guarantees. *Artificial Intelligence*, 111(1–2):209–238, 1999.

[15] O. Shehory and S. Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1-2):165–200, 1998.

[16] M. Wooldridge and P. E. Dunne. On the computational complexity of qualitative coalitional games. *Artificial Intelligence*, 158(1):27–73, 2004.

[17] M. Wooldridge and P. E. Dunne. On the computational complexity of coalitional resource games. *Artificial Intelligence*, 170(10):853–871, 2006.