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 "$\perp$" (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 $\perp$) 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 $\perp$ 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 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. □

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 $\gamma_{a_1} = \Psi \wedge (\neg d) \vee (\wedge_{v\in\Phi} v)$. □

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 $\Phi \notin [\![\gamma_i]\!]$ 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 $\Phi \in [\![\gamma_i]\!]$. 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 $\Phi \notin [\![\gamma_i]\!]$ 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 $\Phi \setminus \{x\} \in [\![\gamma_i]\!]$ 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:

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

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:

$$\exists x \forall y[(x \vee y) \wedge (x \vee \neg y)] \quad (2)$$

This formula is in fact clearly true, as witnessed for example by the existential variable assignment $\{x\}$. Let $\bar{x} = \{x_1, \dots, x_g\}$ be the universally quantified variables in the input formula, let $\bar{y} = \{y_1, \dots, y_h\}$ 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: $A = \{1, \dots, g + h\}$;
- 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 $2 \le i \le g$, fix $\Phi_i = \{x_i\}$;
- for all $1 \le i \le g$ fix $\gamma_i = \chi(\bar{x}, \bar{y}) \wedge d$
- for all $g + 1 \le i \le g + h$, 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 $\zeta_1, \dots, \zeta_6$, 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 $core(G) \subseteq [\![\varphi]\!]$?

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 $[\![\varphi]\!] \subseteq core(G)$?

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 $(\xi \models \varphi)$?