A Modal Logic for Coalitional Power in Games

MARC PAULY, *Center for Mathematics and Computer Science (CWI), P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.*
E-mail: pauly@cwi.nl

Abstract

We present a modal logic for reasoning about what groups of agents can bring about by collective action. Given a set of states, we introduce *game frames* which associate with every state a strategic game among the agents. Game frames are essentially extensive games of perfect information with simultaneous actions, where every action profile is associated with a new state, the outcome of the game. A coalition of players is effective for a set of states X in a game if the coalition can guarantee the outcome of the game to lie in X. We propose a modal logic (*Coalition Logic*) to formalize reasoning about effectivity in game frames, where $[C]\varphi$ expresses that coalition C is effective for φ. An axiomatization is presented and completeness proved. Coalition Logic provides a unifying game-theoretic view of modal logic: Since nondeterministic processes and extensive games without parallel moves emerge as particular instances of game frames, normal and non-normal modal logics correspond to 1- and 2-player versions of Coalition Logic. The satisfiability problem for Coalition Logic is shown to be PSPACE-complete.

Keywords: Modal logic, game theory, multiagent systems.

1 Introduction

Modelling actions and their effects is a task which has occupied many researchers in computer science, logic, economics and philosophy. In the simplest case, we have one agent (person, process) who can choose between taking different actions which change the state of the world in various ways. A simple model of this scenario will contain an accessibility relation R which associates to every state of the world all those states which the agent can bring about through his actions, i.e. sRt holds if in state s the agent can act so as to bring about state t. In modal logic, one introduces a language to talk about such Kripke models: $\Diamond\varphi$ expresses that the agent can act in such a way that φ will be true after his action.

This simple one-agent case can easily be extended to many agents by considering a relational structure which contains an accessibility relation R_i for every agent i, where $\Diamond_i\varphi$ expresses that agent i can bring about φ. The problem with such a multi-agent action logic is that it considers the different agents in isolation. Given a state s_0, agent 1 may act to bring about state s_1 and agent 2 may act to bring about state s_2, but what happens if both of them act simultaneously in s_0? Since the actions of the two agents will often not be independent but interact with each other, a more general model of action should associate a resulting state with every pair of actions (a_1, a_2) of the two players rather than with actions of the players individually.

In this paper, we develop a modal logic based on such more general action models which we shall call *game frames*. At any state of such a frame, each agent $i \in N$ takes an action, and taken together these actions determine the resulting state. This amounts to associating a strategic game form with every state of the frame where the outcomes of the game are states of the frame again. Thus, game frames are essentially extensive game forms with simultaneous actions (see [9]).

In Section 2, game frames are introduced together with extensive games without simultaneous moves as well as non-deterministic processes as special cases. Section 3 relates a notion of effectivity to strategic games, formalizing what it means for a coalition of agents to have the ability to force a certain set of outcomes in a strategic game. This notion of effectivity will then be used as the basic semantic notion for the modal logic we develop in Section 4. For a set of agents $C \subseteq N$, the modal language will contain formulas $[C]\varphi$ which express that the group of agents C can bring about φ, i.e. is effective for φ. We provide a complete axiomatization of this logic in Section 5, together with some coalitional principles which serve to restrict the power of coalitions enough to yield an axiomatization of extensive games without simultaneous moves. Section 6 discusses the complexity of the satisfiability problem for coalition logic. The possibility for agents to combine strategies when forming a coalition is responsible for making this problem PSPACE-complete rather than NP-complete. Finally, Section 7 provides a unifying game-theoretic view of modal logic where normal as well as non-normal modal logics emerge as restricted versions of Coalition Logic.

The logic introduced here can be viewed as a generalization of the modal base logic underlying Parikh's game logic GL [10, 11], an extension of Propositional Dynamic Logic. GL is a logic of determined 2-player games, though a multi-player version is also discussed. The generalization of Coalition Logic consists of dropping the assumption of determinacy and extending the language from individual players to groups of players. While operations on games are not the concern of this paper, such operations could also be added to Coalition Logic, see the remarks in Section 8.

2 A model of interaction: game frames

As mentioned in the introduction, we would like an action model where at each state, the actions taken by the agents together determine the resulting state. To obtain such a model, we associate a strategic game with every state of the world. A *strategic game* $G = (N, \{\Sigma_i | i \in N\}, o, S)$ consists of a non-empty finite set of agents or players N, a non-empty set of strategies or actions Σ_i for every player $i \in N$, a non-empty set of outcome states S and an outcome function $o : \Pi_{i \in N}\Sigma_i \to S$ which associates with every tuple of strategies of the players (strategy profile) an outcome state in S.

In game theory [9, 2], strategic games also come equipped with a preference relation $\succeq_i \subseteq S \times S$ for every player $i \in N$ which indicates which outcomes a player prefers. Strictly speaking, our strategic games are only game *forms* which can be turned into a game by adding these preference relations.

For notational convenience, let $\sigma_C := (\sigma_i)_{i \in C}$ denote the strategy tuple for coalition $C \subseteq N$ which consists of player i choosing strategy $\sigma_i \in \Sigma_i$. Then given two strategy tuples σ_C and $\sigma_{\overline{C}}$ (where $\overline{C} := N \setminus C$), $o(\sigma_C, \sigma_{\overline{C}})$ denotes the outcome state associated with the strategy profile induced by σ_C and $\sigma_{\overline{C}}$.

Let Γ_S^N be the set of all strategic games among the set of players N over the set of states S. Then we define a *game frame* for players N as a pair (S, γ) where S is a non-empty set of states and $\gamma : S \to \Gamma_S^N$ is a function which associates strategic games to states. In game theoretic terminology, game frames are essentially extensive game forms with simultaneous moves [9], the only difference being that we assume that at every state some game can be played, i.e. there are no terminal positions in the game. This assumption that γ is a total function is a matter of convenience, since it allows for a smoother comparison with traditional modal logics (Section 7). Furthermore, nothing is lost making this assumption, since terminal positions can always be equipped with a dummy game and a special label indicating that it is terminal. In any case, results about axiomatization and complexity can be generalized to the case where γ is partial.

Game frames are models of interaction which generalize other well-known action models. First, one may require that agents do not act in parallel but only consecutively. This is captured by extensive game forms without simultaneous moves, and can be modelled e.g. by a standard Kripke models with one accessibility relation which links states to successor states and propositional letters being used to indicate which player has to move. These extensive games without simultaneous moves can also be characterized within the class of game frames. Call a strategic game $G = (N, \{\Sigma_i | i \in N\}, o, S)$ a *d-dictatorship* iff

$$\forall s \in \text{ range}(o)\ \exists \sigma_d \forall \sigma_{N \setminus \{d\}}\ \ o(\sigma_d, \sigma_{N \setminus \{d\}}) = s.$$

Note that in case there is more than one dictator, the outcome function is constant (i.e. $\exists s \forall \sigma\ o(\sigma) = s$) and hence every player is a dictator. A *dictatorial frame* is a game frame (S, γ) such that for every $s \in S$ there is some $d \in N$ such that $\gamma(s)$ is a d-dictatorship. It should be clear that dictatorial frames correspond to the Kripke models just mentioned: the dictator is the one who can choose between the different successor states.

Non-deterministic processes are an even more restricted class of game frames: there is only one player, and consequently this player can decide at every state what the successor state should be. In this case, no interaction arises and we can do with a standard Kripke frame to model such processes. In terms of game frames, a *process* is a game frame (S, γ) for a set of players N where $|N| = 1$. It is easy to see that every process is also a dictatorial frame, where the same dictator is in power at every state.

3 Strategic games and effectiveness

In the next section, we propose a logic for describing and reasoning about the interaction model described in the previous section. The property we want to reason about is coalitional effectiveness: assuming a particular state s, is a coalition of players $C \subseteq N$ effective in achieving a set of states $X \subseteq S$? The notion of effectivity to be employed comes from social choice theory where it goes under the name of α-effectivity [7, 1].

Given a game G, a coalition $C \subseteq N$ will be *effective* for a set $X \subseteq S$ iff the coalition has a joint strategy which will result in an outcome in X no matter what strategies the other players choose. Formally, the *effectivity function* $E_G : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *of a game* G is defined as

$$X \in E_G(C) \text{ iff } \exists \sigma_C \forall \sigma_{\overline{C}}\ \ o(\sigma_C, \sigma_{\overline{C}}) \in X.$$

As the quantifier combination suggests, effectivity in strategic games has both existential and universal character. This is mirrored in our informal translation of $X \in E_G(C)$ as 'coalition C can force X'.

Besides looking at effectiveness in strategic games, the notion of effectiveness can also be investigated more generally. For a set of agents N and a set of outcomes S, we call any function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ an *effectivity function*. Intuitively, E associates to every coalition $C \subseteq N$ the sets of outcomes for which the coalition is effective, i.e. $X \in E(C)$ iff C is effective for X.

Given this intended interpretation of effectivity functions, it is natural to require them to satisfy certain properties, e.g. if a coalition C is effective for a set of outcomes X, it should also be effective for $X' \supseteq X$. Which properties we require will depend on the exact situation we want to model. Call an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *outcome-monotonic* iff for all $X \subseteq X' \subseteq S$ and for all C, if $X \in E(C)$ then $X' \in E(C)$. E is *coalition-monotonic* iff for $C \subseteq C' \subseteq N$, $E(C) \subseteq E(C')$. E is C-*regular* if for all X, if $X \in E(C)$ then $\overline{X} \notin E(\overline{C})$. E is C-*maximal* if for all X, if $\overline{X} \notin E(\overline{C})$ then $X \in E(C)$. E is *regular* (*maximal*) iff for all coalitions C it is C-regular (C-maximal). Finally, E is *superadditive* if for all X_1, X_2, C_1, C_2 such that $C_1 \cap C_2 = \emptyset$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$ imply that $X_1 \cap X_2 \in E(C_1 \cup C_2)$.

Effectivity functions have been investigated in [8, 1, 12] and find application in the theory of social choice [7] where the agents are voters who try to force certain election outcomes. While we consider any function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ to be an effectivity function, most of the literature has taken a more restrictive view, requiring E to satisfy various basic properties, some of which (e.g. monotonicity) have been mentioned. The choice of these basic properties, however, is somewhat arbitrary, and this opinion is supported by the fact that authors differ in which basic properties they require. Here, we decided to be as general as possible regarding the notion of an effectivity function, codifying all the properties we are interested in in the notion of playability to be introduced in the next section.

3.1 Characterizing effectivity in strategic games

The question to be examined here is which effectivity functions are effectivity functions of some strategic game. As we will see in the next section, this characterization result will be very useful when formulating the semantics of coalition logic, since it allows us to dispense with strategic games by only talking about effectivity functions. Two such characterization results have been obtained in [7, 12]. However, these assume that the outcome function o of a strategic game is surjective, whereas we do not want to assume that every state can be reached provided that the players pick the right strategies; certain states may be unreachable no matter how the players play. For the same reason, we did not assume that playable effectivity functions satisfy $X \in E(N)$ for all $X \neq \emptyset$; usually, this is considered to be an essential property of effectivity functions [7, 12]. Thus, while the proof of Theorem 3.2 below makes use of the techniques applied in [7, 12], it generalizes the results obtained previously.

We now introduce the combination of properties needed to characterize effectivity functions. Call an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *playable* iff (1) $\forall C \subseteq N : \ \emptyset \notin E(C)$, (2) $\forall C \subseteq N : \ S \in E(C)$, (3) E is N-maximal, (4) E is outcome-monotonic, and (5) E is superadditive. It can be shown that these conditions are independent.

LEMMA 3.1
Every playable effectivity function is regular and coalition-monotonic.

PROOF. For regularity, let $X \in E(C)$ and assume by reductio that $\overline{X} \in E(\overline{C})$. By superadditivity, $\emptyset \in E(N)$, contradicting condition (1) of playability.

For coalition monotonicity, let $X \in E(C)$ and $C \subseteq C'$. For $C'' := C' \setminus C$, condition (2) of playability give us $S \in E(C'')$ and so by superadditivity, $X \in E(C \cup C'') = E(C')$. ∎

THEOREM 3.2 (Characterization)
An effectivity function E is playable iff it is the effectivity function of some strategic game.

PROOF. One can easily check that the effectivity function of any strategic game satisfies the five properties mentioned. As for the other direction, let E be an effectivity function satisfying the five properties. We shall construct a game G such that $E = E_G$. To simplify our definitions, assume that $N = \{1, \ldots, n\}$.

To guide the reader through the following technical proof, we first provide a more informal sketch of the main argument: given the playable effectivity function E, we construct a strategic game $G = (N, \{\Sigma_i | i \in N\}, o, S)$. A strategy σ_i for player i will be a triple (f_i, t_i, h_i): for every coalition C_i of which i is a member, the function f_i picks a set which C_i can force, and for every non-empty set X, the function h_i picks an element of X. Thus, if player i ends up as a member of coalition C_i, he will force $f_i(C_i)$, and if the choice is up to him, he will pick the outcome $h_i(f_i(C_i))$. Since all players will force certain sets as part of their strategy σ_i, we use the t_is to determine which player will get the power to determine the outcome state. The outcome of the game will be determined by the outcome function o roughly as follows: given $(f_1, \ldots, f_n)$, N is partitioned into coalitions (as big as possible) such that all members of a coalition chose to force the same set. The outcome set will then be the result of each coalition forcing their set, i.e. the intersection of all the sets forced. The player who chooses which state in this set will be realized is then determined by adding up (modulo n) all the indices chosen as t_i. The effectivity function of this game is just E.

Formally, for $i \in N$, let $\mathcal{C}_i = \{C \subseteq N | i \in C\}$ be the set of coalitions of which i is a member. Let $F_i = \{f_i : \mathcal{C}_i \to \mathcal{P}(S) \mid \forall C : f_i(C) \in E(C)\}$, so F_i consist of all functions f_i which associate to every coalition C in which i participates a set of outcomes for which C is effective. Note that since for all coalitions C, $S \in E(C)$, F_i will be non-empty for every player i.

Given $f \in \Pi_{i \in N} F_i = F_N$ and a coalition C, let $P(f, C)$ be the coarsest partition $\langle C_1, \ldots, C_m \rangle$ of C such that

$$\forall l \leq m \forall i, j \in C_l : f_i(C) = f_j(C).$$

Then given f, let

$$\begin{array}{rcl} P_0(f) & = & \langle N \rangle \\ P_1(f) & = & P(f, N) = \langle C_1^1, \ldots, C_{k_1}^1 \rangle \\ P_2(f) & = & \langle P(f, C_1^1), \ldots, P(f, C_{k_1}^1) \rangle = \langle C_1^2, \ldots, C_{k_2}^2 \rangle \\ & \vdots & \\ P_r(f) & = & \langle C_1^r, \ldots, C_{k_r}^r \rangle. \end{array}$$

Since there are only finitely many players, this partitioning process will eventually stop at some state r where $P_r(f) = P_{r+1}(f)$, and we let $P_\infty(f) = P_r(f) = \langle C_1, \ldots, C_k \rangle$. Since for all $l \leq k$ and $i, j \in C_l$ we have $f_i(C_l) = f_j(C_l)$ we will simply write $f(C_l)$ for it. Now let

$$G(f) = \bigcap_{l=1}^{k} f(C_l).$$

Claim: $G(f) \neq \emptyset$. Proof: Since C_l is effective for $f(C_l)$, i.e. $f(C_l) \in E(C_l)$ for all $l \leq k$, $\bigcap_{l=1}^{k} f(C_l) = G(f) \in E(N)$ by superadditivity, and hence since $\emptyset \notin E(N)$, $G(f)$ cannot be empty.

Now we can define the strategic game $G = (N, \{\Sigma_i | i \in N\}, o, S)$ as follows: let $H = \{h : \mathcal{P}(S) \setminus \{\emptyset\} \to S \mid h(X) \in X\}$. Then we define $\Sigma_i = F_i \times N \times H$ and $o(\sigma_N) = h_{i_0}(G(f))$, where $\sigma_N = (f_i, t_i, h_i)_{i \in N}$ is a strategy profile and $i_0 = ((t_1 + \cdots + t_n) \mod n) + 1$

indicates the player who has the power to determine the outcome. It remains to show that for all $C \subseteq N$, $E(C) = E_G(C)$.

For the inclusion from left to right, assume that $X \in E(C)$. Choose any C-strategy $\sigma_C = (f_i, t_i, h_i)_{i \in C}$ such that for all $i \in C$ and for all $C' \supseteq C$ we have $f_i(C') = X$. By coalition-monotonicity, such f_i exist. Take any $\overline{C}$-strategy $\sigma_{\overline{C}} = (f_i, t_i, h_i)_{i \in \overline{C}}$. We need to show that $o(\sigma_C, \sigma_{\overline{C}}) \in X$. To see this, note that C must be a subset of one of the partitions C_l in $P_\infty(f)$. Hence,

$$o(\sigma_N) = o(\sigma_C, \sigma_{\overline{C}}) = h_{i_0}(G(f)) \in G(f) \subseteq f(C_l) = X.$$

For the inclusion from right to left, assume that $X \notin E(C)$. Suppose first that $C = N$. Then by N-maximality, $\overline{X} \in E(\emptyset)$, and by the previous part of the proof, $\overline{X} \in E_G(\emptyset)$. Since E_G is playable, it is regular (by the previous lemma) and so $X \notin E_G(N)$, and we have established the result.

So assume from now on that $C \neq N$, and let $j_0 \in N \setminus C$. Let σ_C be any strategy for coalition C. We must show that there is a strategy $\sigma_{\overline{C}}$ such that $o(\sigma_C, \sigma_{\overline{C}}) \notin X$. Define $\sigma_{\overline{C}} = (f_i, t_i, h_i)_{i \in \overline{C}}$ such that for all $C' \supseteq \overline{C}$ and for all $i \in \overline{C}$ we have $f_i(C') = S$. Then choose a t_{j_0} such that $((t_1 + \cdots + t_n) \bmod n) + 1 = j_0$.

Note that $\overline{C}$ must be a subset of one of the partitions C_l in $P_\infty(f)$. For the other partitions, superadditivity implies that there is some $C_0 \subseteq C$ such that $G(f) \in E(C_0)$, and hence by coalition-monotonicity, $G(f) \in E(C)$. Since $X \notin E(C)$, $G(f) \nsubseteq X$ by outcome-monotonicity, so there is some $s_0 \in G(f) \cap \overline{X}$. Now we define $h_{j_0}(G(f)) = s_0$. Then

$$o(\sigma_C, \sigma_{\overline{C}}) = h_{j_0}(G(f)) = s_0 \notin X.$$

∎

3.2 *Characterizing effectivity in dictatorships*

Besides characterizing the effectivity functions of strategic games in general, we can also try to characterize the effectivity functions of particular subclasses of strategic games, in particular dictatorships, given their relevance for defining extensive games without simultaneous moves.

Call an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *individualistic* iff it is playable and $E(N) = \bigcup_{i \in N} E(\{i\})$. The condition ensures that everything which can be forced can be forced already by some individual. The following result shows that individualism is an extremely strong assumption: while it seems to say only that the whole is equal to the sum of its parts, it actually says that the whole is equal to one particular part.

THEOREM 3.3
An effectivity function E is individualistic iff it is the effectivity function of a dictatorship.

PROOF. First, if E is the effectivity function of a dictatorship with dictator d, E is easily seen to be individualistic.

Second, assume E is individualistic, and so there is a strategic game G such that $E = E_G$. We can assume that G has at least two distinct outcomes t_1 and t_2, for otherwise G is trivially a dictatorship. Suppose by reductio that there are two individuals $i \neq j \in N$ such that $\{t_1\} \in E(\{i\})$ and $\{t_2\} \in E(\{j\})$. Then by superadditivity, $\emptyset \in E(\{i, j\})$, a contradiction. Hence, there is a player who can force any outcome, so G is a dictatorship. ∎

Put positively, unless we have a dictatorship, coalitions of agents can sometimes achieve more than their members individually, cooperation is thus advantageous.

3.3 Semi-playability and N-maximality

In some of the definitions used in the completeness and complexity arguments to follow we will define effectivity functions E with two separate clauses, one for all coalitions $C \neq N$, and another one for $C = N$. Since the objective is to obtain playable effectivity functions, the definition will make sure that $X \in E(N)$ iff $\overline{X} \notin E(\emptyset)$. Verifying that an effectivity function defined in such a way is playable is facilitated by the following lemma. It makes use of the notion of semi-playability, to be thought of as playability for all coalitions other than N, excluding the condition of N-maximality.

Call an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ *semi-playable* iff (1) $\forall C \neq N$: $\emptyset \notin E(C)$, (2) $\forall C \neq N$: $S \in E(C)$, (3) E is outcome-monotonic for all coalitions $C \neq N$, and (4) E is superadditive for all coalitions $C \neq N$, i.e. for all X_1, X_2, C_1, C_2 such that $C_1 \cap C_2 = \emptyset$ and $C_1 \cup C_2 \neq N$, $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$ imply that $X_1 \cap X_2 \in E(C_1 \cup C_2)$. Together with N-maximality and regularity, semi-playability suffices for playability.

LEMMA 3.4
An effectivity function E is playable iff it is semi-playable, regular and N-maximal.

PROOF. All the playability conditions for N follow almost immediately from the corresponding semi-playability conditions for $\emptyset$ together with maximality or regularity. We only show superadditivity: assume $X_1 \in E(C_1)$ and $X_2 \in E(C_2)$ where $C_1 \cap C_2 = \emptyset$ and $C_1 \cup C_2 = N$. Consider the case where $C_1 = N$, $C_2 = \emptyset$. Assume by reductio that $X_1 \cap X_2 \notin E(C_1 \cup C_2)$. By N-maximality, $\overline{X_1 \cap X_2} \in E(\emptyset)$ and by superadditivity (given by semi-playability), $\overline{X_1} \cap X_2 \in E(\emptyset)$. By monotonicity, $\overline{X_1} \in E(\emptyset)$ and by regularity, $X_1 \notin E(N)$, a contradiction. Other cases where $C_1 = \overline{C_2}$ are treated analogously. ∎

4 Syntax and semantics of coalition logic

Given a finite non-empty set of agents/players N, we define the syntax of Coalition Logic as follows. Given a set of atomic propositions Φ_0, a formula φ can have the following syntactic form:

$$\varphi := \quad \bot \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid [C]\varphi$$

where $p \in \Phi_0$ and $C \subseteq N$. We define $\top$, $\wedge$, $\to$ and $\leftrightarrow$ as usual. In the case $C = \{i\}$, we we write $[i]\varphi$ instead of $[\{i\}]\varphi$.

A *coalition frame* is a pair $\mathcal{F} = (S, E)$ where S is a non-empty set of states (the universe) and

$$E : S \to (\mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S)))$$

is the *playable effectivity structure* of the model: for every state $s \in S$, $E(s)$ is a playable effectivity function. For easier readability, we shall often write $sE_C X$ instead of $X \in E(s)(C)$ to denote that C is effective for X at state s.

A *coalition model* is a pair $\mathcal{M} = (\mathcal{F}, V)$ where $\mathcal{F}$ is a coalition frame and $V : \Phi_0 \to \mathcal{P}(S)$ is the usual valuation function for the propositional letters.

From what was said in the preceding section, our restriction to playable effectivity functions should be clear: Theorem 3.2 guarantees that coalition frames are just game frames.

Since $E(s)$ is the effectivity function of a strategic game, every state s is associated with a strategic game $G(s)$, and $sE_C X$ holds iff coalition C is effective for X in $G(s)$.

Given such a model, truth of a formula in a model at a state is defined as follows

$$\begin{array}{lll} \mathcal{M}, s \not\models \bot & & \\ \mathcal{M}, s \models p & \text{iff} & p \in \Phi_0 \text{ and } s \in V(p) \\ \mathcal{M}, s \models \neg\varphi & \text{iff} & \mathcal{M}, s \not\models \varphi \\ \mathcal{M}, s \models \varphi \vee \psi & \text{iff} & \mathcal{M}, s \models \varphi \text{ or } \mathcal{M}, s \models \psi \\ \mathcal{M}, s \models [C]\varphi & \text{iff} & sE_C\varphi^{\mathcal{M}} \end{array}$$

where $\varphi^{\mathcal{M}} = \{s \in S | \mathcal{M}, s \models \varphi\}$. Hence, a formula $[C]\varphi$ holds at a state s iff coalition C is effective for $\varphi^{\mathcal{M}}$ in $G(s)$. A formula φ is *valid in a model* $\mathcal{M}$ with universe S, denoted as $\mathcal{M} \models \varphi$, iff $\varphi^{\mathcal{M}} = S$, and φ is *valid* in a class of models K (denoted as $\models_{\mathsf{K}} \varphi$) iff for all models $\mathcal{M} \in \mathsf{K}$ we have $\mathcal{M} \models \varphi$. We write $\Sigma \models_{\mathsf{K}} \varphi$ to denote that φ is a (local) logical consequence of Σ: for all models $\mathcal{M} \in \mathsf{K}$ and every state s of the universe of $\mathcal{M}$, if $\mathcal{M}, s \models \Sigma$ (i.e. all formulas of Σ are true at s) then $\mathcal{M}, s \models \varphi$.

Let M be the class of all coalition models, and let $\mathsf{M_d}$ be the class of coalition models $((S, E), V)$ where for all $s \in S$, $E(s)$ is individualistic. Given the characterization result for dictatorship from the previous section, we can thus think of $\mathsf{M_d}$ as the class of dictatorial game models.

Thanks to theorem 3.2, we are able to represent coalition models without referring to games and strategies. This simplifies the meta-theoretic treatment of our logic, and it also demonstrates that a coalition model is simply a multi-modal generalization of a neighbourhood model (or minimal model, see Section 7.2 and [4]), providing a neighbourhood relation for every coalition of players. Neighbourhood models have been the standard semantic tool to investigate non-normal modal logics, and techniques used to provide complete axiomatization for such logics can also be adapted to Coalition Logic.

5 Axiomatization

Given a set of players N, a *coalition logic for* N is a set of formulas Λ which contains all propositional tautologies together with all instances of the axiom schemas listed in Figure 1, and which is closed under the rules of Modus Ponens and Equivalence:

$$\frac{\varphi \quad \varphi \to \psi}{\psi} \qquad\qquad \frac{\varphi \leftrightarrow \psi}{[C]\varphi \leftrightarrow [C]\psi}$$

Let $\mathtt{CL_N}$ denote the smallest coalition logic for N. Notice that the axioms are direct translations of the five playability conditions into the modal language, outcome monotonicity being captured by axiom $\mathtt{M}$.

Given a coalition logic Λ, we write $\vdash_\Lambda \varphi$ for $\varphi \in \Lambda$ and $\Sigma \vdash_\Lambda \varphi$ if there exist $\sigma_1, \ldots, \sigma_n \in \Sigma$ such that $(\sigma_1 \wedge \ldots \wedge \sigma_n) \to \varphi \in \Lambda$. Finally, a set of formulas Σ is Λ-inconsistent iff $\Sigma \vdash_\Lambda \bot$.

LEMMA 5.1
The Monotonicity rule $\dfrac{\varphi \to \psi}{[C]\varphi \to [C]\psi}$ is derivable in any coalition logic Λ.

PROOF. Let $\vdash_\Lambda \varphi \to \psi$. By propositional logic, $\vdash_\Lambda (\varphi \wedge \psi) \leftrightarrow \varphi$ and by the equivalence rule $\vdash_\Lambda [C](\varphi \wedge \psi) \leftrightarrow [C]\varphi$. The monotonicity axiom $\mathtt{M}$ together with propositional logic then yield $\vdash_\Lambda [C]\varphi \to [C]\psi$. ∎

$(\bot)$	$\neg[C]\bot$
$(\top)$	$[C]\top$
(N)	$(\neg[\emptyset]\neg\varphi \to [N]\varphi)$
(M)	$[C](\varphi \wedge \psi) \to [C]\psi$
(S)	$([C_1]\varphi_1 \wedge [C_2]\varphi_2) \to [C_1 \cup C_2](\varphi_1 \wedge \varphi_2)$
	where $C_1 \cap C_2 = \emptyset$

FIGURE 1. The axiom schemas of coalition logic

Call a logic Λ *complete* with respect to a class of coalition models K if $\Sigma \models_{\mathsf{K}} \varphi$ iff $\Sigma \vdash_\Lambda \varphi$. The rest of this section is devoted to showing that $\mathtt{CL}_\mathtt{N}$ is complete for the class of all coalition models. The proof is via a canonical model construction.

Let Λ be any coalition logic. Via the standard argument of Lindenbaum's lemma, every Λ-consistent set of formulas Σ can be extended to a maximally Λ-consistent set $\Sigma' \supseteq \Sigma$ with the usual properties: (1) for every formula φ, $\varphi \in \Sigma'$ or $\neg\varphi \in \Sigma'$, (2) $\varphi \vee \psi \in \Sigma'$ iff $\varphi \in \Sigma'$ or $\psi \in \Sigma'$, and (3) if $\Sigma' \vdash_\Lambda \varphi$ then $\varphi \in \Sigma'$.

Now let S^Λ be the set of all maximally Λ-consistent sets of formulas, and let $\tilde{\varphi} := \{s \in S^\Lambda | \varphi \in s\}$. Define the canonical Λ-model $\mathcal{C}^\Lambda = ((S^\Lambda, E^\Lambda), V^\Lambda)$ as follows:

$$\begin{array}{rl} s \in V^\Lambda(p) & \text{iff } p \in s; \\ sE^\Lambda_C X & \text{iff } \left\{ \begin{array}{ll} \exists \tilde{\varphi} \subseteq X : [C]\varphi \in s & \text{for } C \neq N, \\ \forall \tilde{\varphi} \subseteq \overline{X} : [\emptyset]\varphi \notin s & \text{for } C = N. \end{array} \right. \end{array}$$

To see that E^Λ is well-defined, note that if $\tilde{\varphi_1} = \tilde{\varphi_2}$, $\vdash_\Lambda \varphi_1 \leftrightarrow \varphi_2$ and so $\vdash_\Lambda [C]\varphi_1 \leftrightarrow [C]\varphi_2$ which implies that for all $s \in S^\Lambda$, $[C]\varphi_1 \in s$ iff $[C]\varphi_2 \in s$. Note also that we defined $sE^\Lambda_N X$ iff not $sE^\Lambda_\emptyset \overline{X}$. To check that we have indeed defined a coalition model, we need to check that the playability conditions are met.

LEMMA 5.2
For all $s \in S^\Lambda$, $E^\Lambda(s)$ is playable.

PROOF. By Lemma 3.4, it is sufficient to check that for any maximally Λ-consistent set $s \in S^\Lambda$, $E^\Lambda(s)$ is semi-playable.

(1) Let $C \neq N$ and assume by reductio that there is some $\tilde{\varphi} \subseteq \emptyset$ such that $[C]\varphi \in s$. Then $\tilde{\varphi} = \tilde{\bot}$ and so $[C]\bot \in s$, a contradiction. (2) For $C \neq N$, since $[C]\top \in s$ and $\tilde{\top} \subseteq S^\Lambda$, $sE^\Lambda_C S^\Lambda$. (3) Monotonicity is easily seen to hold by definition. (4) For $C_1, C_2 \neq N$ and $C_1 \cap C_2 = \emptyset$, assume that $[C_1]\varphi_1, [C_2]\varphi_2 \in s$ for $\tilde{\varphi_1} \subseteq X_1$ and $\tilde{\varphi_2} \subseteq X_2$. By the superadditivity axiom, $[C_1 \cup C_2](\varphi_1 \wedge \varphi_2) \in s$, and since $\varphi_1 \tilde{\wedge} \varphi_2 \subseteq X_1 \cap X_2$, we have $sE^\Lambda_{C_1 \cup C_2}(X_1 \cap X_2)$. ∎

LEMMA 5.3 (Truth Lemma)
For any maximally Λ-consistent set $s \in S^\Lambda$ and any formula φ: $\mathcal{C}^\Lambda, s \models \varphi$ iff $\varphi \in s$. Equivalently, $\varphi^{\mathcal{C}^\Lambda} = \tilde{\varphi}$.

PROOF. For atomic formulas and for the Boolean inductive steps, the argument is standard. For $[C]\varphi$, it is sufficient to consider the case where $C \neq N$. So suppose $s \in [C]\varphi^{\mathcal{C}^\Lambda}$, i.e. there is some $\tilde{\varphi_0} \subseteq \varphi^{\mathcal{C}^\Lambda}$ such that $[C]\varphi_0 \in s$. Since by induction hypothesis $\varphi^{\mathcal{C}^\Lambda} = \tilde{\varphi}$, $\vdash_\Lambda \varphi_o \to \varphi$ and so using the derived monotonicity rule, $[C]\varphi \in s$ as well.

Conversely, if $[C]\varphi \in s$, given that $\varphi^{\mathcal{C}^\Lambda} = \tilde{\varphi}$ by induction hypothesis, the result follows immediately. ■

THEOREM 5.4 (Canonical Model Theorem)
Every coalition logic Λ is complete with respect to its canonical model $\mathcal{C}^\Lambda$.

PROOF. Let Λ be any coalition logic. One can easily show by induction on the length of a derivation that $\Sigma \vdash_\Lambda \varphi$ implies $\Sigma \models_{\{C^\Lambda\}} \varphi$. For the converse, suppose $\Sigma \not\vdash_\Lambda \varphi$, so $\Sigma \cup \{\neg\varphi\}$ is Λ-consistent, and so there is a maximally Λ-consistent set $\Sigma' \in S^\Lambda$ with $\Sigma \cup \{\neg\varphi\} \subseteq \Sigma'$ such that $\mathcal{C}^\Lambda, \Sigma' \models \Sigma$ while $\mathcal{C}^\Lambda, \Sigma' \not\models \varphi$, showing that $\Sigma \not\models_{\{C^\Lambda\}} \varphi$. ■

Using Lemma 5.2 and the canonical model theorem, we obtain the following completeness result.

THEOREM 5.5 (Completeness)
$\mathtt{CL_N}$ is complete with respect to the class of all coalition models: $\Sigma \models_{\mathsf{M}} \varphi$ iff $\Sigma \vdash_{\mathtt{CL_N}} \varphi$.

Axiomatizing dictatorship

The logic $\mathtt{CL_N}$ is the most general and hence weakest coalition logic. The only assumption made is that at every state, the coalitional power distribution arises from a situation which can be modelled as a strategic game. The stronger assumptions made for dictatorial frames also lead to an axiomatizable logic. Let $\mathtt{DCL_N}$ be the smallest coalition logic including the two axioms

$$(\mathtt{D_1}) \quad [N]\varphi \to \textstyle\bigvee_{i\in N}[i]\varphi,$$
$$(\mathtt{D_2}) \quad [N]\neg\varphi \wedge [N]\psi \to \textstyle\bigwedge_{i\in N}([i]\varphi \to [i]\psi).$$

Let Λ be any coalition logic extending $\mathtt{DCL_N}$. We again consider the set of maximally Λ-consistent sets of formulas S^Λ. The crucial lemma needed for the completeness proof is that every $s \in S^\Lambda$ has a local dictator.

LEMMA 5.6
For any $s \in S^\Lambda$, there is some $d \in N$ such that for all formulas φ, if $[N]\varphi \in s$ then $[d]\varphi \in s$.

PROOF. Suppose by reductio that there is some $s \in S^\Lambda$ such that for every $i \in N$ there is a formula φ with $[N]\varphi \in s$ and $[i]\varphi \notin s$. This implies that $|N| > 1$, and that there are two players $i \neq j \in N$ and two formulas φ and ψ such that $[N]\varphi, [N]\psi, [j]\varphi \in s$ (we have used axiom $\mathtt{D_1}$), while $[i]\varphi, [j]\psi \notin s$. We also know that $[N]\neg\varphi \in s$, for otherwise $\neg[N]\neg\varphi \in s$ and so by axiom $\mathtt{N}$, $[\emptyset]\varphi \in s$ and using axiom $\mathtt{S}$, $[i]\varphi \in s$, a contradiction. But since $[N]\neg\varphi \wedge [N]\psi \in s$, axiom $\mathtt{D_2}$ gives $[j]\varphi \to [j]\psi \in s$, and hence $[j]\psi \in s$, a contradiction. ■

As remarked earlier, the presence of a (local) dictator essentially turns coalition models into Kripke models. The completeness proof below should be seen as a translation of the completeness proof for a normal modal logic into the coalitional setup.

THEOREM 5.7
$\Sigma \vdash_{\mathtt{DCL_N}} \varphi$ iff $\Sigma \models_{\mathsf{M_d}} \varphi$.

PROOF. For any coalition logic $\Lambda \supseteq \mathtt{DCL_N}$, we construct the canonical model $\mathcal{C}^\Lambda$ as before, except that for E^Λ, we let

$$sE^\Lambda_C X \quad \text{iff} \begin{cases} \exists x \in X\, \forall\varphi \in x\; [C]\varphi \in s & \text{for } d_s \in C \\ \forall x \notin X\, \exists\varphi \in x\; [\overline{C}]\varphi \notin s & \text{for } d_s \notin C, \end{cases}$$

where d_s is the local dictator of s provided by the previous lemma. It can be verified that for all $s \in S^\Lambda$, $E(s)$ is individualistic. Furthermore, note that $E(s)$ is not only N-maximal but maximal, and that for every C with $d_s \in C$, $sE^\Lambda_{\overline{C}}X$ implies $sE^\Lambda_C X$.

Next, one can establish the truth lemma for $\mathcal{C}^\Lambda$ which makes use of the standard existence lemma: for all $s \in S^\Lambda$ with $[C]\varphi \in s$ and $d_s \in C$, there is some $x \in \tilde{\varphi}$ such that for all $\delta \in x$, $[C]\delta \in s$. We construct this x by defining $x_0 = \{\varphi\} \cup \{\delta | [\overline{C}]\delta \in s\}$. It can be shown that x_0 is consistent, so one can construct a maximally consistent set $x \subseteq x_0$ which satisfies the conditions of the existence lemma. The proof continues as before. ∎

6 Complexity

6.1 Upper bound

We will show that for a formula φ of length n, there is a deterministic algorithm requiring space polynomial in n which computes whether or not φ is satisfiable, i.e. whether or not $\vdash_{\mathtt{CL_N}} \neg\varphi$. The heart of the algorithm relies on Lemma 6.1 which reduces the satisfiability of φ to the satisfiability of certain combinations of subformulas of φ which have smaller modal depth. This method of providing a PSPACE-algorithm is adapted from [14], where complexity results for various non-normal modal logics are proved.

Let $sf(\varphi)$ be the set of subformulas of φ, let

$$X_\varphi = sf(\varphi) \cup \{[N]\neg\delta | [\emptyset]\delta \in sf(\varphi)\} \cup \{[\emptyset]\neg\delta | [N]\delta \in sf(\varphi)\} \cup \{[C]\top, [C]\bot | C \subseteq N\} \cup \{\bot\}$$

and we can set $Cl(\varphi) = X_\varphi \cup \{\neg\delta | \delta \in X_\varphi\}$. Note that $Cl(\varphi)$ is finite and that it is still closed under subformulas and their negations. A *semi-valuation* for φ is a function $v : Cl(\varphi) \to \{0,1\}$ such that (1) $v(\psi) = 1$ iff $v(\neg\psi) = 0$, (2) $v(\psi_1 \vee \psi_2) = 1$ iff $v(\psi_1) = 1$ or $v(\psi_2) = 1$, (3) $v(\bot) = 0$, and (4) $v(\varphi) = 1$.

The following lemma provides the crucial link between satisfiability of a formula and satisfiability of its subformulas. Intuitively, the four conditions of the following lemma mirror the five playability requirements, where the last condition captures both superadditivity and monotonicity.

LEMMA 6.1
A formula φ is satisfiable iff there exists a semi-valuation v for φ such that the following four conditions hold:

1. If $[C_1]\psi_1, \ldots, [C_k]\psi_k \in Cl(\varphi)$, $\forall m \in N$: $m \in C_i \cap C_j \Rightarrow i = j$, and $\forall i :\ v([C_i]\psi_i) = 1$, then $\bigwedge_i \psi_i$ is satisfiable.
2. If $[C]\psi \in Cl(\varphi)$ and $v([C]\psi) = 0$, then $\neg\psi$ is satisfiable.
3. If $[\emptyset]\psi_1, [N]\psi_2 \in Cl(\varphi)$ and $v([\emptyset]\psi_1) = v([N]\psi_2) = 0$, then $\neg\psi_1 \wedge \neg\psi_2$ is satisfiable.
4. If $[C]\psi, [C_1]\psi_1, \ldots, [C_k]\psi_k \in Cl(\varphi)$, $\forall m \in N$: $m \in C_i \cap C_j \Rightarrow i = j$, $C = \bigcup_i C_i$, $v([C]\psi) = 0$ and $\forall i :\ v([C_i]\psi_i) = 1$, then $\neg\psi \wedge \bigwedge_i \psi_i$ is satisfiable.

PROOF. From left to right, suppose φ is satisfiable in a coalition model $\mathcal{M} = ((S,E),V)$ at state $s \in S$. Then v defined by $v(\psi) = 1$ iff $\mathcal{M}, s \models \psi$ is a semi-valuation for φ, and it will satisfy the four conditions by virtue of $E(s)$ being playable. Suppose $v([\emptyset]\psi_1) = v([N]\psi_2) = 0$, i.e. $\mathcal{M}, s \models \neg[\emptyset]\psi_1 \wedge \neg[N]\psi_2$. By N-maximality, $\mathcal{M}, s \models [N]\neg\psi_1$ and hence $\neg\psi_1^{\mathcal{M}} \not\subseteq \psi_2^{\mathcal{M}}$, i.e. $\neg\psi_1 \wedge \neg\psi_2$ must be satisfiable. Similarly for the other conditions.

From right to left, suppose we have a semi-valuation v satisfying the four conditions. This means that for every $[C]\psi \in Cl(\varphi)$ for which $v([C]\psi) = 0$, there is a model $\mathcal{M}$ and a state s such that $\mathcal{M}, s \models \neg\psi$, and similarly for the other conditions. Thus, we have a sequence of models $\mathcal{M}_1, \ldots, \mathcal{M}_n$ and a sequence of states $s_1, \ldots, s_n$ which serve as witnesses to the four conditions. We can assume w.l.o.g. that the universes of these models are pairwise disjoint, i.e. for all $\mathcal{M}_i = ((S_i, E_i), V_i)$ and $\mathcal{M}_j = ((S_j, E_j), V_j)$ with $i \neq j$ we have $S_i \cap S_j = \emptyset$. To simplify notation, we shall also use $V_i(\psi)$ for $\psi^{\mathcal{M}_i}$ when ψ is not atomic.

We will now construct a model $\mathcal{M} = ((S, E), V)$ satisfying φ which is roughly the union of the $\mathcal{M}_i$ models. For a new state s_0 which shall correspond to v, let $S = \{s_0\} \cup \bigcup_{i>0} S_i$. Let $V_0 : Cl(\varphi) \to \mathcal{P}(\{s_0\})$ be defined as $V_0(\psi) = \{s_0\}$ if $v(\psi) = 1$ and $\emptyset$ otherwise. Let $J : Cl(\varphi) \to \mathcal{P}(S)$ be defined by $J(\psi) = \bigcup_{i\geq 0} V_i(\psi)$. By construction, $J(\neg\psi) = S \setminus J(\psi)$ and $J(\psi_1 \vee \psi_2) = J(\psi_1) \cup J(\psi_2)$.

To complete the definition of our newly constructed model, let $V(p) = J(p)$ for $p \in \Phi_0$. For $s_a \neq s_0$ and $C \neq N$, let $E(s_a)$ be defined as follows:

$$\begin{array}{ll} s_a E_C X \text{ iff} & \exists [C_1]\psi_1, \ldots, [C_k]\psi_k \in Cl(\varphi): \\ & (1)\ C = \bigcup_i C_i,\ (2)\ \forall m \in N : m \in C_i \cap C_j \Rightarrow i = j \\ & (3)\ \bigcap_i J(\psi_i) \subseteq X, \text{ and } (4)\ \mathcal{M}_a, s_a \models \bigwedge_i [C_i]\psi_i. \end{array}$$

Note that in the above, k may be 1, in which case the right-hand side reduces to $\exists [C]\psi \in Cl(\varphi)$ such that $J(\psi) \subseteq X$ and $\mathcal{M}_a, s_a \models [C]\psi$. For the set of all players, we define $s_a E_N X$ iff not $s_a E_\emptyset \overline{X}$.

For s_0, we use an analogous definition in terms of the semi-valuation v: for $C \neq N$, $E(s_0)$ is defined as above, except that we replace condition (4) by $\forall i : v([C_i]\psi_i) = 1$. Again, we define $s_0 E_N X$ iff not $s_0 E_\emptyset \overline{X}$.

$\mathcal{M}$ is semi-playable at every state s_a. For $s_a \neq s_0$, the playability conditions either hold by definition or they are essentially inherited from $\mathcal{M}_a$. For s_0, the first two conditions of the lemma are used.

Claim 1: If $C \neq N$, $[C]\psi \in Cl(\varphi)$, and $s_a E_C J(\psi)$ for $a > 0$, then $\mathcal{M}_a, s_a \models [C]\psi$. The proof uses superadditivity and monotonicity of $\mathcal{M}_a$.

Claim 2: If $C \neq N$, $[C]\psi \in Cl(\varphi)$, and $s_0 E_C J(\psi)$, then $v([C]\psi) = 1$. Suppose that $\exists [C_1]\psi_1, \ldots, [C_k]\psi_k \in Cl(\varphi)$ such that $\bigcap_i J(\psi_i) \subseteq J(\psi)$, and assume by reductio that $v([C]\psi) = 0$ while $\forall i : v([C_i]\psi_i) = 1$. Then using condition 4 of the lemma, $\neg\psi \wedge \bigwedge_i \psi_i$ is satisfiable, and so for some a we must have $\mathcal{M}_a, s_a \models \neg\psi \wedge \bigwedge_i \psi_i$, contradicting the fact that $\bigcap_i J(\psi_i) \subseteq J(\psi)$.

To show that $\mathcal{M}, s_0 \models \varphi$, we show that for all $\psi \in Cl(\varphi)$, $V(\psi) = J(\psi)$. The proof is by induction on ψ, and base case and Boolean cases are immediate. For $C \neq N$, let $s_a \in J([C]\psi)$. Depending on a, this either means that $\mathcal{M}_a, s_a \models [C]\psi$ or that $v([C]\psi) = 1$. In both cases, $s_a E_C J(\psi)$ holds and by induction hypothesis, $s_a E_C V(\psi)$ and hence $s_a \in V([C]\psi)$. Similarly for the converse direction, using claim 1 and 2.

For $C = N$, condition 3 of the lemma is used. Suppose $s_a \in V([N]\psi)$, i.e. $s_a \notin V([\emptyset]\neg\psi)$, and hence by the previous argument $s_a \notin J([\emptyset]\neg\psi)$. In the case that $s_a \neq s_0$, $\mathcal{M}_a, s_a \not\models [\emptyset]\neg\psi$, and N-maximality gives $s_a \in J([N]\psi)$. In the case $s_a = s_0$, $v([\emptyset]\neg\psi) = 0$. Now if we assume by reductio that $v([N]\psi) = 0$, condition 3 of the lemma would make $\psi \wedge \neg\psi$ satisfiable, a contradiction, and so $v([N]\psi) = 1$, establishing $s_0 \in J([N]\psi)$. The other direction makes use of condition 1. ∎

THEOREM 6.2

The satisfiability problem for Coalition Logic is in PSPACE.

PROOF. Consider the following description of the satisfiability game for formula φ:

Game Sat(φ)

(1) construct $Cl(\varphi)$

(2) $\exists$-player: choose a semi-valuation v for φ

(3) if φ contains no modalities, $\exists$-player wins; otherwise:

(4) $\forall$-player: choose a condition (1-4);

(5) if condition = 1 then

(5.1) $\forall$-player: choose a subset $[C_1]\psi_1, \ldots, [C_k]\psi_k \in CL(\varphi)$ such that the coalitions are pairwise disjoint, and for all i: $v([C_i]\psi_i) = 1$

(5.2) continue playing Sat($\bigwedge_i \psi_i$)

(6) if condition = 2 then

(6.1) $\forall$-player: choose $[C]\psi \in CL(\varphi)$ such that $v([C]\psi) = 0$

(6.2) continue playing Sat($\neg\psi$)

(7) if condition = 3 then

(7.1) $\forall$-player: chooses $[\emptyset]\psi_1, [N]\psi_2 \in CL(\varphi)$ such that $v([\emptyset]\psi_1) = v([N]\psi_2) = 0$

(7.2) continue playing Sat($\neg\psi_1 \wedge \neg\psi_2$)

(8) if condition = 4 then

(8.1) $\forall$-player: choose a subset $[C]\psi, [C_1]\psi_1, \ldots, [C_k]\psi_k \in CL(\varphi)$ such that the coalitions C_i are pairwise disjoint, $C = \bigcup_i C_i$, and for all i: $v([C_i]\psi_i) = 1$ and $v([C]\psi) = 0$

(8.2) continue playing Sat($\neg\psi \wedge \bigwedge_i \psi_i$).

Assuming that a player who cannot choose as instructed loses (e.g. $\exists$-player loses in step 2 if there is no semi-valuation for φ), we have defined a 2-player game. By Lemma 6.1, $\exists$-player has a winning strategy in this game iff φ is satisfiable. To analyse the time it takes to play the game, i.e. the maximal length of a play, let n be the length of formula φ, or equivalently $n = |sf(\varphi)|$. For the purposes of this algorithm, we allow for generalized conjunctions $\bigwedge_i \psi_i$ where $sf(\bigwedge_i \psi_i) = \{\bigwedge_i \psi_i\} \cup \bigcup_i sf(\psi_i)$. The construction of $Cl(\varphi)$ and checking whether v is a semi-valuation takes time linear in n, and the size of $Cl(\varphi)$ is also linear in n. For condition 1, a maximum of n formulas $[C_i]\psi_i$ are chosen so that checking all $v([C_i]\psi_i)$ takes time $O(n^2)$. Finally, the game is continued with $\bigwedge_i \psi_i$ whose size is $|sf(\bigwedge_i \psi_i)| \leq |sf(\varphi)| = n$. Note that this recursive call reduces the modal depth of the formula by 1 until eventually φ contains no more modalities, hence the number of recursive calls is at most n.

Thus, at most n rounds of Sat are played, each round taking time polynomial in n. Since the size of each game configuration is also polynomial in n, doing backward induction on the game tree can be done in PSPACE by a depth-first search algorithm. In other words, since Sat(φ) contains a high-level description of an Alternating Turing Machine [3] there is an alternating polynomial time algorithm for satisfiability checking, and given that APTIME = PSPACE, this means that there is a deterministic polynomial space algorithm for satisfiability. ∎

6.2 Lower bound

To show that the satisfiability problem for Coalition Logic is also PSPACE-hard, we show that the normal modal logic $\mathtt{KD}$ is a sublogic of Coalition Logic, and hence its satisfiability problem (which is PSPACE-hard) can be reduced to satisfiability in Coalition Logic.

Consider the one-player coalition logic $\mathtt{CL}_{\{1\}}$. Denoting $[\emptyset]\varphi$ as $\Box\varphi$ and $[\{1\}]\varphi$ as $\Diamond\varphi$, $\mathtt{CL}_{\{1\}}$ is the smallest set of formulas containing all propositional tautologies and closed under Modus Ponens and Equivalence rule which contains the axioms of Figure 2.

$\neg\Box\bot$	$\neg\Diamond\bot$
$\Box\top$	$\Diamond\top$
$\neg\Box\neg\varphi \to \Diamond\varphi$	
$\Box(\varphi\wedge\psi) \to \Box\psi$	$\Diamond(\varphi\wedge\psi) \to \Diamond\psi$
$\Box\varphi\wedge\Box\psi \to \Box(\varphi\wedge\psi)$	$\Diamond\varphi\wedge\Box\psi \to \Diamond(\varphi\wedge\psi)$

FIGURE 2. Axioms of $\mathtt{CL}_{\{1\}}$

$\mathtt{KD}$ is the normal modal logic for reasoning about serial Kripke models, i.e. Kripke models where every state has at least one successor [4]. In the formulation closest to coalition logic, $\mathtt{KD}$ is the set of formulas containing all propositional tautologies, closed under the rules of Modus Ponens and Equivalence (for $\Box$ only), and containing the axioms of Figure 3.

$\Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$	$\Box(\varphi\wedge\psi) \leftrightarrow (\Box\varphi\wedge\Box\psi)$
$\Box\top$	$\Diamond\top$

FIGURE 3. Axioms of $\mathtt{KD}$

THEOREM 6.3
The satisfiability problem for Coalition Logic is PSPACE-hard.

PROOF. First, it can easily be shown that $\mathtt{KD} = \mathtt{CL}_{\{1\}}$ by induction on the length of a derivation. Second, if N is any non-empty set of players, for every formula φ of $\mathtt{CL}_{\{1\}}$, $\varphi \in \mathtt{CL}_{\{1\}}$ iff $\varphi^\circ \in \mathtt{CL}_{\mathtt{N}}$, where φ° is the same as φ except that coalition N is substituted for coalition $\{1\}$. Inspecting the axioms, one sees that $\varphi \in \mathtt{CL}_{\{1\}}$ implies that $\varphi^\circ \in \mathtt{CL}_{\mathtt{N}}$. For the other direction, if $\varphi \notin \mathtt{CL}_{\{1\}}$, there is a coalition model $\mathcal{M}_1$ satisfying $\neg\varphi$. It suffices to observe that $\mathcal{M}_1$ can easily be turned into a coalition model $\mathcal{M}_N$ for the set of players N which satisfies $\neg\varphi^\circ$.

As a result, there is a polynomial time translation from φ into a formula φ° such that φ is valid in $\mathtt{KD}$ iff φ° is valid in Coalition Logic, where the length of φ° is polynomial in the length of φ. Hence the satisfiability problem of $\mathtt{KD}$ is polynomial time reducible to the satisfiability problem of Coalition Logic. Since $\mathtt{K} \subseteq \mathtt{KD} \subseteq \mathtt{S4}$, by Ladner's theorem [6], the satisfiability problem for $\mathtt{KD}$ is PSPACE-hard, and by the reduction, the satisfiability problem of Coalition Logic is PSPACE-hard as well. ∎

6.3 The complexity of coalition formation

When investigating various non-normal epistemic logics in [14], the author observes a complexity difference which hinges on the presence of the formula C

$$K\varphi \wedge K\psi \rightarrow K(\varphi \wedge \psi),$$

where $K\varphi$ should be read as 'the agent knows that φ'. Among the various epistemic systems investigated, logics which do not contain this principles have their satisfiability problem in NP, whereas those containing C have it in PSPACE. While in the latter case, no lower bound is proved, it is conjectured that this principle formalizes an agent's ability to epistemically combine facts, i.e. to reason about the world, and that it is this ability which causes the (conjectured) complexity increase.

Assuming that NP $\neq$ PSPACE, an analogue of this conjecture can be proved for Coalition Logic. As can be seen from the form that the superadditivity axiom S takes in $\mathtt{CL}_{\{1\}}$ (Figure 2), superadditivity is the game-theoretic analogue of the epistemic principle C. Instead of expressing the ability of an agent to combine facts, it expresses the ability of agents to combine their strategies when forming a coalition. And in the case of Coalition Logic, it is possible to locate a complexity increase precisely in this ability to combine strategies, since one can show that for Coalition Logic without superadditivity, the satisfiability problem is NP-complete.

Without going into much detail, let $\mathtt{CL}_N^-$ be Coalition Logic without the superadditivity axiom S, and assume that coalition models consist of effectivity functions which may or may not be superadditive. Since propositional logic is a part of $\mathtt{CL}_N^-$, the satisfiability problem of $\mathtt{CL}_N^-$ is NP-hard. For a non-deterministic polynomial time algorithm solving the satisfiability problem, we again make use of a modified version of Lemma 6.1, where condition (1) is replaced by

if $[C]\psi \in Cl(\varphi)$ and $v([C]\psi) = 1$, then ψ is satisfiable;

condition (4) is replaced by simple monotonicity:

if $[C]\psi_1, [C]\psi_2 \in Cl(\varphi)$, $v([C]\psi_1) = 1$ and $v([C]\psi_2) = 0$, then $(\psi_1 \wedge \neg\psi_2)$ is satisfiable.

and the definition of E is given by $s_a E_C X$ iff there is some $[C]\psi \in Cl(\varphi)$ such that $J(\psi) \subseteq X$ and $\mathcal{M}_a, s_a \models [C]\psi$ (or $v([C]\psi) = 1$, in the case $a = 0$). The crucial difference with the original lemma is that for $\mathtt{CL}_N^-$, the number of recursive satisfiability checks needed is at most $|Cl(\varphi)|^2$ whereas in the original lemma, any subset of $[C_i]\psi_i$ formulas could be chosen, requiring at most $2^{|Cl(\varphi)|}$ satisfiability checks.

Using dynamic programming techniques, we can determine the satisfiability of all formulas of the form $\psi_1 \wedge \psi_2$, where $\psi_i \in CL(\varphi)$, in polynomial time. Hence, the algorithm runs in non-deterministic polynomial time, showing that the satisfiability problem for $\mathtt{CL}_N^-$ is NP-complete. Again assuming that NP $\neq$ PSPACE, this demonstrates that reasoning about systems where agents are able to form coalitions is more complex than reasoning about systems where there is no coalition formation.

7 A game-theoretic view of modal logic

7.1 One-player games: normal modal logic

In establishing a lower complexity bound we already uncovered the close connection between $\mathtt{CL}_{\{1\}}$ and normal modal logic:

THEOREM 7.1
Identifying $[\emptyset]\varphi$ with $\Box\varphi$ and $[\{1\}]\varphi$ with $\Diamond\varphi$, we have $\mathtt{KD} = \mathtt{CL}_{\{1\}}$.

If the set of players is a singleton such as $\{1\}$, coalition models correspond to game frames which are processes. Since every state is associated with a game in which an output state can be chosen, these coalition models are serial Kripke frames $\mathcal{M} = ((S, R), V)$ where $R \subseteq S \times S$ is the accessibility relation linking states to output/successor states and truth of a modal formula is defined by

$$\mathcal{M}, s \models \Box\varphi \text{ iff } \forall x : \; sRx \Rightarrow \mathcal{M}, x \models \varphi. \tag{7.1}$$

Intuitively, non-serial Kripke frames equally well correspond to processes, the question arises however how coalitional power should be defined at states where no game can be played. It would seem natural to postulate that at these gameless states, no coalition can force anything, so $[C]\varphi$ would be false at such states independent of C and φ. While this approach is intuitively appealing, it destroys the $\mathtt{CL}_{\{1\}} = \mathtt{KD}$ equality since at gameless states, $\Box\varphi$ would be true (in normal modal logic) while $[\emptyset]\varphi$ would be false (in Coalition Logic).

7.2 Two-player games: non-normal modal logic

Certain non-normal modal logics also emerge as special cases of Coalition Logic. Non-normal modal logics [4] describe *neighbourhood models* $\mathcal{M} = ((S, N), V)$ which are almost like coalition models except that they contain only a single effectivity function, i.e. $N : S \to \mathcal{P}(\mathcal{P}(S))$. $\Box\varphi$ will be true at s if there is a neighbourhood of s such that every state in that neighbourhood makes φ true:

$$\mathcal{M}, s \models \Box\varphi \text{ iff } \{t \in S | \mathcal{M}, t \models \varphi\} \in N(s). \tag{7.2}$$

We assume here that for all $s \in S$, $\mathcal{M}$ is monotonic ($X \in N(s)$ implies $X' \in N(s)$ provided that $X \subseteq X'$), $S \in N(s)$ and $\emptyset \notin N(s)$.

It can be shown that for these neighbourhood models, the set of valid formulas $\mathtt{MD}$ is the set of formulas containing all propositional tautologies, closed under the rules of Modus Ponens and Equivalence (for $\Box$), and containing the axioms of figure 4.

$\Diamond\varphi \leftrightarrow \neg\Box\neg\varphi$	$\Box(\varphi \wedge \psi) \to \Box\psi$
$\Box\top$	$\Diamond\top$

FIGURE 4. Axioms of $\mathtt{MD}$

Consider now the logic $\mathtt{DetCL}_{\{1,2\}}$, the smallest coalition logic for $\{1, 2\}$ containing the following axiom of determinacy:

$$\neg[\overline{C}]\neg\varphi \to [C]\varphi.$$

The axiom expresses that for every set of states X, either coalition C is effective for X or coalition $\overline{C}$ is effective for $\overline{X}$.

The following theorem shows that the logic $\mathtt{MD}$ is nothing but deterministic 2-player Coalition Logic restricted to formulas talking about singleton coalitions.

THEOREM 7.2
Identifying $[1]\varphi$ with $\Box\varphi$ and $[2]\varphi$ with $\Diamond\varphi$, we have

$$\mathtt{MD} = \mathtt{DetCL}_{\{1,2\}} \cap \{\varphi | \text{if } C \text{ occurs in } \varphi \text{ then } |C| = 1\}.$$

PROOF. $\mathtt{MD} \subseteq \mathtt{DetCL}_{\{1,2\}}$: The only non-obvious case is showing that $[2]\varphi \to \neg[1]\neg\varphi \in \mathtt{DetCL}_{\{1,2\}}$ which follows from superadditivity. Conversely, for every formula φ containing only the two singleton coalitions, $\varphi \in \mathtt{DetCL}_{\{1,2\}}$ implies $\varphi \in \mathtt{MD}$. Given a neighbourhood model $\mathcal{M} = ((S,N),V)$ satisfying $\neg\varphi$, construct a full coalition model $\mathcal{M}' = ((S,E),V)$ by defining $sE_{\{1\}}X$ iff $X \in N(s)$, $sE_{\{2\}}X$ iff $\overline{X} \notin N(s)$, $sE_{\{1,2\}}X$ iff there are X_1, X_2 with $X = X_1 \cap X_2$, $sE_{\{1\}}X_1$ and $sE_{\{2\}}X_2$. Finally, $sE_{\emptyset}X$ iff not $sE_{\{1,2\}}\overline{X}$. All the axioms of $\mathtt{DetCL}_{\{1,2\}}$ are valid in $\mathcal{M}'$ and it will satisfy $\neg\varphi$; hence, $\varphi \notin \mathtt{DetCL}_{\{1,2\}}$. ∎

7.3 n-player games: Coalition Logic

Moving from 2 to n players we move from non-normal modal logic to general Coalition Logic. One special case that has been discussed for n-player games was the class of dictatorial frames, i.e. extensive games without simultaneous moves, with its associated Coalition Logic $\mathtt{DCL_N}$. Figure 5 presents an overview of the systems discussed so far, where $|N| = n$.

Type of games	Modal logic	Coalition Logic
1 player	normal, KD	$\mathtt{CL}_{\{1\}}$
2 players, determined	non-normal, MD	$\mathtt{DetCL}_{\{1,2\}}$
n players, no simultaneous moves	-	$\mathtt{DCL_N}$
n players	-	$\mathtt{CL_N}$

FIGURE 5. Overview of modal logics and coalition logics

While we have come a long way from normal modal logic, note that possibility and necessity are still a part of general Coalition Logic: Game-theoretically, the possibility of φ is the existence of an outcome where φ holds and is thus expressed by $[N]\varphi$, whereas the necessity of φ corresponds to φ holding at every outcome of the game, $[\emptyset]\varphi$. Hence, possibility and necessity form extreme cases of coalitional power.

8 Extensions and applications

The aim of the present work has been largely foundational. The scope of modal logic has been extended to cover reasoning about coalitional effectivity in extensive games with simultaneous actions. The effectivity modality itself turns out to be an interesting hybrid between possibility and necessity. In the process, a result from social choice theory (Theorem 3.2) played a central role in obtaining a complete axiomatization. On the other hand, the perspective of modal logic led to a new dynamic view of effectivity (coalition frames) which in turn produced a generalization of the original characterization results obtained in [7, 12].

The present paper has already investigated two instances of additional axioms which lay down the relationship between what a group and what its members can force, namely weak and strong dictatorship. As has been shown in Section 3.2, even such very simple principles may have surprising consequences. Nonetheless, the present logic allows for investigating the consequences of less trivial relationships between group and group-member effectivity, e.g. *democratic effectivity*, where $[N]p \rightarrow [C]p$ holds provided $|C| > \frac{1}{2}|N|$.

A further extension of the present work is the addition of a modality for expressing *global* rather than local effectivity. While $[C]\varphi$ refers to C's immediate effectiveness in bringing about φ, one may also want to reason about what the coalition can bring about in the extensive game as a whole, i.e. what terminal positions can be forced. This requires adding an iterated modality $[C^*]\varphi$ in the style of dynamic logic [5], and such an extension would allow for investigating the interplay between local and global coalitional principles. The result would be a step towards a coalitional multi-player version of Parikh's game logic.

Finally, the present approach ignores players' preferences altogether. The logic only provides means for reasoning about what coalitions *can* bring about, not what they *will* bring about. Especially when one also adds a belief or knowledge structure to the models, adding preferences and solution concepts could lead to a logic of coalition formation, formalizing some of the processes described in [13].

Acknowledgements

I thank Johan van Benthem, Peter van Emde Boas, Rohit Parikh, Yde Venema and the anonymous referees for comments on earlier drafts of this paper.

References

[1] J. Abdou and H. Keiding. *Effectivity Functions in Social Choice*. Kluwer, 1991.

[2] K. Binmore. *Fun and Games – A Text on Game Theory*. Heath, 1992.

[3] A. Chandra, D. Kozen, and L. Stockmeyer. Alternation. *Journal of the ACM*, **28**, 114–133, 1981.

[4] B. Chellas. *Modal Logic - An Introduction*. Cambridge University Press, 1980.

[5] D. Harel. Dynamic logic. In *Handbook of Philosophical Logic*, volume II, D. Gabbay and F. Guenthner, eds. Reidel, 1984.

[6] R. Ladner. The computational complexity of provability in systems of modal propositional logic. *SIAM Journal of Computing*, **6**, 467–480, 1977.

[7] H. Moulin. *The Strategy of Social Choice*. North-Holland, 1983.

[8] H. Moulin and B. Peleg. Cores of effectivity functions and implementation theory. *Journal of Mathematical Economics*, **10**, 115–145, 1982.

[9] M. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[10] R. Parikh. Propositional logics of programs: New directions. In *Foundations of Computation Theory*, volume 158 of *LNCS*, M. Karpinski, ed. pp. 347–359. Springer, 1983.

[11] R. Parikh. The logic of games and its applications. In *Topics in the Theory of Computation*, M. Karpinski and J. van Leeuwen, eds. Volume 24 of *Annals of Discrete Mathematics*. Elsevier, 1985.

[12] B. Peleg. Effectivity functions, game forms, games and rights. *Social Choice and Welfare*, **15**, 167–80, 1998.

[13] F. Tohmé and T. Sandholm. Coalition formation processes with belief revision among bounded-rational self-interested agents. *Journal of Logic and Computation*, **9**, 793–815, 1999.

[14] M. Vardi. On the complexity of epistemic reasoning. In *Proceedings of the 4th annual Symposium on Logic in Computer Science*, pp. 243–252, 1989.

Received 2 May 2000