Reasoning about the Power of Coalitions

Marc Pauly

CWI, P.O. Box 94079, 1090 GB Amsterdam. E-mail: pauly@cwi.nl

Social Choice Theory (see e.g. [3]) has developed a general model for coalitional power which allows one to analyze e.g. whether or to what extent a group of voters can force a certain alternative to be chosen. The central notion employed is that of an *effectivity function* (see [5]): Given a finite set of agents N and a set of alternatives A, an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(A))$ associates with every coalition of agents the sets of alternatives which the coalition can force, i.e. $X \in E(C)$ if the coalition C can force the alternative chosen to lie in X.

The concept of an effectivity function is static in the sense that it specifies sets of abstract alternatives which coalitions can bring about irrespective of the situation at hand. By taking the view of possible-worlds semantics, we relativize effectivity functions to states of the world. Secondly, we take the alternatives to be possible worlds again, yielding a dynamic model of coalitional power in which coalitions of agents can restrict the navigation through the state space in various ways.

Formally, the language of *N-agent Coalition Logic* is a multi-modal logic where the modalities are indexed by coalitions $C \subseteq N$, and formulas φ are of the form

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

where $p \in \Phi_0$ is an atomic proposition. $[C]\varphi$ expresses that coalition C is effective for φ. A coalition model for the set of agents N is a triple $\mathcal{M} = (S, E, V)$ where S is a nonempty set of states (the universe), $V : \Phi_0 \to \mathcal{P}(S)$ is the usual valuation function for the propositional letters, and

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

is the *global effectivity structure* of the model: For every state $s \in S$, $E(s)$ is an effectivity function. Coalition models are essentially minimal models with one neighborhood relation (see [2]) for every coalition.

Depending on the type of interaction the agents are involved in, different requirements will have to be imposed on the global effectivity structure of the models. To give one example, it seems reasonable in most situations to assume that if a coalition C is effective for X, any larger coalition $C' \supseteq C$ will also be effective for X. Yet, it could be the case that a new member of the larger coalition can undermine the effectiveness of some members of C, so that C' turns out to be less effective than C.

A very general model of multi-agent interaction is that of a strategic game (see [6]). A play of such a game $G = (N, \{\Sigma_i | i \in N\}, o, S)$ consists of each player $i \in N$ choosing an action (or strategy) $a_i \in \Sigma_i$. The outcome of the game is then determined by the function $o : \Pi_{i \in N} \Sigma_i \to S$ which associates with every action profile an outcome state $s \in S$. Given such a strategic game, a coalition C is said to be α-*effective* for a set of outcomes $X \subseteq S$ iff C has a joint strategy such that for all joint strategies of the remaining players $\overline{C}$, the outcome will be in X. This notion gives rise to an effectivity function E_G^α which can be associated to every strategic game G.

If we want to use our coalition logic to reason about α-effectivity in strategic games, we need to first characterize which effectivity functions are α-effectivity functions of some strategic game. Generalizing results from [4, 7], one can show that an effectivity function $E : \mathcal{P}(N) \to \mathcal{P}(\mathcal{P}(S))$ is the α-effectivity function of a strategic game iff it satisfies five *playability* conditions, the central one being *superadditivity*: 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 $X_1 \cap X_2 \in E(C_1 \cup C_2)$. Imposing these playability conditions on the global effectivity structure of the coalition models, we can see every state as linked to a strategic game which can be played at that state, yielding a new state depending on the strategies or actions chosen by the players. Thus, playable coalition models are general action models, where transitions to successor states are not determined by the agents individually, but by the actions of all the agents together.

The playability conditions can easily be translated into modal formulas which can serve as the axioms of a deductive calculus for Coalition Logic. Due to the fact that effectivity functions are essentially modal neighborhood relations, one can prove soundness and completeness of this axiomatization via an adapted version of the standard canonical model construction.

References

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

[2] Brian Chellas. *Modal Logic - An Introduction.* Cambridge University Press, 1980.

[3] Jerry Kelly. *Social Choice Theory: An Introduction.* Springer, 1988.

[4] Hervé Moulin. *The Strategy of Social Choice.* North-Holland, 1983.

[5] Hervé Moulin and Bezalel Peleg. Cores of effectivity functions and implementation theory. *Journal of Mathematical Economics*, 10(1):115–145, 1982.

[6] Martin Osborne and Ariel Rubinstein. *A Course in Game Theory.* MIT Press, 1994.

[7] Bezalel Peleg. Effectivity functions, game forms, games and rights. *Social Choice and Welfare*, 15:67–80, 1998.