# Scientia Potentia Est*

Thomas Ågotnes
Dept of Information Science and Media Studies
University of Bergen
PB. 7802, 5020 Bergen
Norway
thomas.agotnes@uib.no

Wiebe van der Hoek
Dept of Computer Science
University of Liverpool
Liverpool L69 7ZF
United Kingdom
wiebe@csc.liv.ac.uk

Michael Wooldridge
Dept of Computer Science
University of Liverpool
Liverpool L69 7ZF
United Kingdom
mjw@liv.ac.uk

## ABSTRACT

In epistemic logic, Kripke structures are used to model the distribution of information in a multi-agent system. In this paper, we present an approach to *quantifying* how much information each particular agent in a system has, or how important the agent is, with respect to some fact represented as a goal formula. It is typically the case that the goal formula is distributed knowledge in the system, but that no individual agent alone knows it. It might be that several different groups of agents can get to know the goal formula together by combining their individual knowledge. By using power indices developed in voting theory, such as the Banzhaf index, we get a measure of how important an agent is in such groups. We analyse the properties of this notion of information-based power in detail, and characterise the corresponding class of voting games. Although we mainly focus on distributed knowledge, we also look at variants of this analysis using other notions of group knowledge. An advantage of our framework is that power indices and other power properties can be expressed in standard epistemic logic. This allows, e.g., standard model checkers to be used to quantitatively analyse the distribution of information in a given Kripke structure.

## Categories and Subject Descriptors

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

## General Terms

Theory

## Keywords

Epistemic logic, power indices, model checking

## 1. INTRODUCTION

Epistemic logic is widely used in the multi-agent systems community to reason about the knowledge and ignorance of agents in terms of the information they possess [5]. In many situations, it would be useful to be able to *quantify* how information is distributed in a system, or to reason about the *relative importance* of the information that different agents have. In general, it is difficult to answer the question of whether an agent has more information than another agent except for in special cases, such as when one agent knows everything another agent knows [15]. In this paper, we quantify the distribution of information in a system in a specific sense satisfying two assumptions. The first is that we are interested in who knows more *about* some given fact. The second is that we are interested in situations where information can be *communicated* between agents, and it is not always possible or desirable to communicate with every other agent in the system.

Consider the following situation. $M$ knows that if sales are up this quarter, the stock price will increase ($p \rightarrow q$). $T$ knows that if the new CEO has signed the contract, the stock price will increase ($r \rightarrow q$). $W$ knows that sales are up this quarter and that the new CEO has signed the contract ($p \wedge r$). Assume that this describes all (relevant) facts that the three agents know. Who knows more? We are here interested in a more specific type of question: who has the most *important* or *valuable* information *about* whether or not the stock price will increase ($q$), *in* a social setting where communication is possible? None of the agents alone knows $q$, but they can *combine* their knowledge to find out that $q$ is in fact true. And here the importance of the knowledge of the three agents differ: $M$ and $W$ can together find out $q$, as can $T$ and $W$. $M$ and $T$ cannot. It can thus be argued that $W$ knows more about $q$ in this social setting, since he can combine his knowledge in several different ways with others' knowledge – and, indeed, it is not hard to see that $W$'s knowledge is *necessary* for any group to be able to find out $q$, unlike that of $M$ or $T$. If it is important for each individual agent to find out $q$, and since no agent already knows $q$, the only possibility is to communicate with someone else; in which case clearly $W$ would be considered the most *important* agent.

In this paper we analyse the relative importance of the knowledge each agent has in a system where information about some fact or objective ($q$ in our example above) is distributed throughout the system. To this end, we employ *power indices* such as the Banzhaf index, known from voting theory. The starting point is a pointed Kripke structure. It is typically the case that the objective is distributed knowledge in the system, but that no individual agent knows it. It might be that several different groups of agents can get to know the objective by combining their knowledge. Our approach measures the importance of an agent in an arbitrary group of agents wrt. deriving the objective. We consider an agent to be powerful, or to have important information, if the probability of changing the distributed knowledge in the group from ignorance to knowledge about the objective by joining some arbitrary group, is high. This concept of *information based power* can, e.g., be used to identify agents that are crucial to the functioning of the multi-agent system.

The question of "who knows more" in epistemic logic has re-

**For also Knowledge itself is Power*; with apologies to Francis Bacon.

**Cite as:** Scientia Potentia Est, Thomas Ågotnes and Wiebe van der Hoek and Michael Wooldridge, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 735-742.

cently been studied in [15]. The notion of information based power we introduce in this paper is a more fine-grained generalisation: if an agent knows more in the sense of [15] then she has a higher power index, but not necessarily the other way around. Solution concepts for coalitional games have recently been used to measure the degree of inconsistency in databases [8]. In [2] power indices are used to analyse the relative importance of agents when in terms of complying or not complying with a *normative system* defined over a Kripke-like structure [12, 1]. However, we are not aware of any approaches using power indices to measure relative importance of agents in terms of their knowledge/information as described by a Kripke structure.

The paper is organised as follows. In the two next sections we briefly review some background material about epistemic logic and power indices that we will use. In Section 4 we define power indices for agents, given a pointed Kripke structure and a goal formula. We give a complete characterisation of the power indices that can be obtained in this way, study their properties in detail, and show how standard epistemic logic can be used to express power properties. Since these power properties can be expressed in epistemic logic, we can also use epistemic logic to reason about agents' *knowledge* about such properties. In Section 5 we study what agents know about the distribution of information-based power in the system. In most of the paper we use distributed knowledge to define power, but in Section 6 we discuss other types of group knowledge as well. We conclude in Section 7.

## 2. EPISTEMIC LOGIC

Assume a finite set of agents $Ag = \{1, \ldots, n\}$ and a countably infinite set of atomic propositions $\Theta$. The language $\mathcal{L}_K$ of the epistemic logic $S5_n$ is defined by the following grammar:

$$\varphi ::= \top \mid p \mid K_i\varphi \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2$$

where $p \in \Theta$ and $i \in Ag$. An *epistemic (Kripke) structure*, $M$, (over $Ag$, $\Theta$) is an $(n+2)$-tuple [5]:

$$M = \langle W, \sim_1, \ldots, \sim_n, \pi \rangle, \quad \text{where}$$

- $W$ is a finite, non-empty set of *states*;
- $\sim_i \subseteq W \times W$ is an *epistemic accessibility relation* for each agent $i \in Ag$, where each $\sim_i$ is an equivalence relation; and
- $\pi : W \rightarrow 2^{\Theta}$ is a Kripke valuation function, which gives the set of primitive propositions satisfied in each state.

Formulae are interpreted in a *pointed structure*, a pair $M, s$, where $M$ is a model and $s$ is a state in $M$, as follows.

- $M, s \models \top$
- $M, s \models p$ iff $p \in \pi(s)$ (where $p \in \Theta$)
- $M, s \models \neg\varphi$ iff $M, s \not\models \varphi$
- $M, s \models \varphi \wedge \psi$ iff $M, s \models \varphi$ and $M, s \models \psi$
- $M, s \models K_i\varphi$ iff for all $t$ such that $s \sim_i t$, $M, t \models \varphi$.

We will make use of extensions of $S5_n$ with *group knowledge*. To this end, when $G \subseteq Ag$, we denote the union of $G$'s accessibility relations by $\sim_G^E$, so $\sim_G^E = (\bigcup_{i \in G} \sim_i)$. We use $\sim_G^C$ to denote the transitive closure of $\sim_G^E$. Finally, $\sim_G^D$ denotes the intersection of $G$'s accessibility relations (cf. [5, p.66–70]). The logics $S5_n^D$, $S5_n^C$ and $S5_n^{CD}$ are obtained as follows. The respective languages, $\mathcal{L}_D$, $\mathcal{L}_C$, and $\mathcal{L}_{CD}$, are obtained by adding the clause $D_G\varphi$, $C_G\varphi$, and both, respectively, where $G \subseteq Ag$, to the definition of $\mathcal{L}_K$. The interpretation of the two group operators:

- $M, s \models D_G\varphi$ iff for all $t$ such that $s \sim_G^D t$, $M, t \models \varphi$
- $M, s \models C_G\varphi$ iff for all $t$ such that $s \sim_G^C t$, $M, t \models \varphi$

We use the same notation for the satisfaction relation for all these logics; it will be clear from context which logic we are working in. As usual, we write $M \models \varphi$ if $M, s \models \varphi$ for all $s$ in $M$, and $\models \varphi$ if $M \models \varphi$ for all $M$; in this latter case, we say that $\varphi$ is *valid*. A formula is *satisfied* in a pointed model if it is true. When $\Phi$ is a set of formulae, $\Phi \models \varphi$, $\Phi$ *entails* $\varphi$, means that any pointed model that satisfies $\Phi$ also satisfies $\varphi$. A formula is *satisfiable* if there exists a pointed model that satisfies it. A formula or set of formulae is satisfiable in a *set* of pointed models if it is satisfied by *at least one* pointed model in that set. The usual propositional abbreviations are used, in addition to $E_G\varphi$ ($G \subseteq Ag$) for $\bigwedge_{i \in G} K_i\varphi$; $\hat{K}_i\varphi$ for $\neg K_i\neg\varphi$; $\hat{D}_G\varphi$ for $\neg D_G\neg\varphi$ and $\hat{C}_G\varphi$ for $\neg C_G\neg\varphi$. We will often abuse notation and write singleton sets of agents $\{i\}$ as $i$.

$E_G\varphi$ means that all individuals in the group $G$ know $\varphi$. $D_G\varphi$ means that $\varphi$ is distributed knowledge among $G$. Roughly speaking, this knowledge would come about if all members of $G$ were to share their information (but see also Section 4.2). $C_G\varphi$, that $\varphi$ is common knowledge in $G$, means that $E_G\varphi \wedge E_GE_G\varphi \wedge E_GE_GE_G\varphi \wedge \ldots$. These concepts of group and individual knowledge are related as follows (with $i \in G$):

$$\models (C_G\varphi \rightarrow E_G\varphi) \wedge (E_G\varphi \rightarrow K_i\varphi) \wedge (K_i\varphi \rightarrow D_G\varphi) \wedge (D_G\varphi \rightarrow \varphi)$$

The above implications express that common knowledge is the strongest property, and truth the weakest. However, since $C_G\varphi$ is such a strong notion, this often means it will only be obtained for 'weak' $\varphi$. Or [5], common knowledge can be paraphrased as what 'any fool knows', while distributed knowledge corresponds to what 'a wise man knows'.

Finally, the *knowledge set* of $G \subseteq Ag$ in $M, s$ is:

$$\mathcal{K}_G(M, s) = \{\varphi \in \mathcal{L}_K : M, s \models K_i\varphi \text{ for some } i \in G\}$$

## 3. COALITIONAL GAMES AND POWER

We briefly review some key concepts from the area of cooperative game theory [10] and the theory of voting power [6] that we will use in the following. A *cooperative* (or *coalitional*) *game* is a pair $\Gamma = \langle Ag, \nu \rangle$, where $Ag = \{1, \ldots, n\}$ is a set of *players*, or *agents*, and $\nu : 2^{Ag} \rightarrow \mathbb{R}$ is the *characteristic function* of the game, which assigns to every set of agents a numeric value, which is conventionally interpreted as the value that this group of agents could obtain if they chose to cooperate. A cooperative game is said to be *simple* if the range of $\nu$ is $\{0, 1\}$; in simple games we say that $G$ are *winning* if $\nu(G) = 1$, while if $\nu(G) = 0$, we say $G$ are *losing*. A simple cooperative game is said to be *monotonic* if $\nu(G) = 1$ implies that $\nu(H) = 1$, whenever $G \subseteq H$. A monotonic simple cooperative game is sometimes called a *simple voting game* [6]. For simple games, a number of *power indices* attempt to characterise in a systematic way the *influence* that a given agent has, by measuring how effective this agent is at turning a losing coalition into a winning coalition [6]. The best-known of these is perhaps the *Banzhaf index* and its relatives, the Banzhaf score and Banzhaf measure [3].

Agent $i$ is said to be a *swing player* for $G$ if $G$ is not winning but $G \cup \{i\}$ is. We define a function $swing(G, i)$ so that this function returns 1 if $i$ is a swing player for $G$, and 0 otherwise, i.e.,

$$swing(G, i) = \begin{cases} 1 & \text{if } \nu(G) = 0 \text{ and } \nu(G \cup \{i\}) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Now, we define the *Banzhaf score* for agent $i$, denoted $\sigma_i$, to be the number of coalitions for which $i$ is a swing player:

$$\sigma_i = \sum_{G \subseteq Ag \setminus \{i\}} swing(G, i). \quad (1)$$

The *Banzhaf measure* $\mu_i$, is the probability that $i$ would be a swing player for a coalition chosen at random from $2^{Ag \setminus \{i\}}$:

$$\mu_i = \frac{\sigma_i}{2^{n-1}} \quad (2)$$

The *Banzhaf index* for a player $i \in Ag$, denoted by $\beta_i$, is the proportion of coalitions for which $i$ is a swing to the total number of swings in the game – thus the Banzhaf index is a measure of relative power, since it takes into account the Banzhaf score of other agents:

$$\beta_i = \frac{\sigma_i}{\sum_{j \in Ag} \sigma_j} \quad (3)$$

Finally, we define the *Shapley-Shubik index*; here the *order* in which agents join a coalition plays a role. Let $P(Ag)$ denote the set of all permutations of $Ag$, with typical members $\varpi, \varpi'$, etc. If $\varpi \in P(Ag)$ and $i \in Ag$, then let $prec(i, \varpi)$ denote the members of $Ag$ that precede $i$ in the ordering $\varpi$. Given this, let $\varsigma_i$ denote the Shapley-Shubik index of $i$, defined as follows:

$$\varsigma_i = \frac{1}{|Ag|!} \sum_{\varpi \in P(Ag)} swing(prec(i, \varpi), i) \quad (4)$$

Thus the Shapley-Shubik index is essentially the Shapley value [10, p.291] applied to simple ($\{0, 1\}$-valued) cooperative games.

We say that a player is a *veto player* if it is included in all winning coalitions, a *dictator* if $\mu_i = 1$, and a *dummy* if $\mu_i = 0$.

# 4. POWER OF DISTRIBUTED KNOWLEDGE

We define the power of agents given a pointed Kripke structure, and an objective specified as a *goal formula*. Intuitively, an agent is maximally powerful if she already knows the goal formula, and is completely powerless if she does not know anything needed in combination with others' knowledge to be able to conclude that the goal formula is true. In between these two extremes are potentially many intermediate levels of power: the more sub-groups the agent can join in order for the group to have shared knowledge of the objective, the more powerful the agent is.

In order to formalise the fact that information about the goal formula is shared in a group, we use the concept of distributed knowledge. We define a simple coalitional game where a coalition is winning iff it has distributed knowledge about the goal formula.

Formally, a *goal structure* is a tuple $S = \langle M, s, \chi \rangle$, where $M, s$ is a pointed model over agents $Ag$ and $\chi \in \mathcal{L}_D$ is a goal formula. Given a goal structure we define the simple game $\langle Ag, \nu_S^D \rangle$:

$$\nu_S^D(G) = \begin{cases} 1 & M, s \models D_G \chi \\ 0 & \text{otherwise.} \end{cases}$$

EXAMPLE 1. *Figure 1 shows a model $M_{MTW}$ of the situation described in the introduction. Observe that $M_{MTW}, s \models K_M(p \to q) \wedge K_T(r \to q) \wedge K_W(p \wedge r)$, and also that these formulae represent "private" knowledge of the respective agents; i.e., we have that $M_{MTW}, s \models \neg K_M(r \to q) \wedge \neg K_M(p \wedge r) \wedge \neg K_T(p \to q) \wedge \neg K_T(p \wedge r) \wedge \neg K_W(p \to q) \wedge \neg K_W(r \to q)$. Furthermore observe that $M_{MTW}, s \models \neg D_x q$ for all $x \in \{M, T, W\}$, and that $M_{MTW}, s \models \neg D_{\{M,T\}} q \wedge D_{\{M,W\}} q \wedge D_{\{T,W\}} q$. We thus get that $M$ is swing for exactly $\{W\}$, that $T$ is swing for exactly $\{W\}$, that $W$ is swing for exactly $\{M\}$, $\{T\}$ and $\{M, T\}$, and thus that:*

$$\sigma_M = \sigma_T = 1, \sigma_W = 3 \quad \mu_M = \mu_T = \tfrac{1}{4}, \mu_W = \tfrac{3}{4}$$
$$\beta_M = \beta_T = \tfrac{1}{5}, \beta_W = \tfrac{3}{5} \quad \varsigma_M = \varsigma_T = \tfrac{1}{6}, \varsigma_W = \tfrac{2}{3}.$$

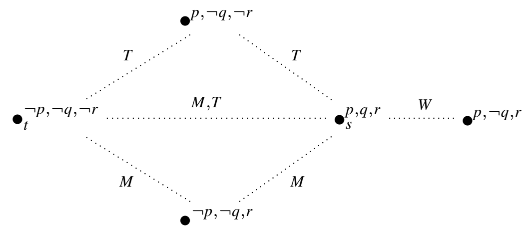


**Figure 1: The model $M_{MTW}$. Reflexive loops are omitted.**

What are the properties of $\nu_S^D$? From the fact that $D_G \chi$ implies $D_H \chi$ when $G \subseteq H$ it follows that $\nu_S^D$ is always *monotonic*. In fact, monotonicity completely characterise the (simple) games induced in this way: every monotonic (voting) game is induced by some Kripke structure and goal formula via the definition above.

THEOREM 1. *For any simple cooperative game $\Gamma = \langle Ag, \nu \rangle$, there exists a goal structure $S$ such that $\nu_S^D = \nu$ iff $\Gamma$ is monotonic.*

PROOF. The implication to the right is immediate (as already mentioned), so assume that $\nu$ is monotonic. Let $p \in \Theta$. We construct a goal structure $S = \langle M, s, \chi \rangle$ such that $\nu_S^D = \nu$ as follows: $W = \{s_0\} \cup \{s_H : \nu(H) = 0\}$; $s = s_0$; $V(p) = \{s_0\}$; $\chi = p$. $\sim_i$ is defined by the following equivalence classes: $[s_0]_{\sim_i} = \{s_0\} \cup \{s_H : i \in H\}$ and for every $H'$ such that $i \notin H'$, $[s_{H'}]_{\sim_i} = \{s_{H'}\}$. Informally: for each $H$ such that $\nu(H) = 0$ there is a designated state $s_H$ where $p$ is false, which no agent in $H$ can discern from $s_0$.

Let $\nu(G) = 1$. We must show that $M, s_0 \models D_G p$, so let $t$ be such that $(s_0, t) \in \bigcap_{i \in G} \sim_i$. It suffices to show that $t = s_0$. Assume otherwise: that $t = s_H$ for some $H$ such that $\nu(H) = 0$. For every $i \in G$, $s_0 \sim_i s_H$, and by the definition of $\sim_i$ it follows that $i \in H$. Thus, $G \subseteq H$. But since $\nu(G) = 1$ and $\nu(H) = 0$, that contradicts monotonicity.

Conversely, let $\nu(G) = 0$. We have that $s_0 \sim_i s_G$ for every $i \in G$ and $M, s_G \models \neg p$. Thus $M, s_0 \not\models D_G p$. $\square$

## 4.1 Expressing Power

Epistemic logic can be used to express and reason about power in Kripke structures. The following expressions can, e.g., be used together with a standard model checker, to determine the power distribution in a given structure.

- $i$ is swing for $G$ when the goal is $\chi$:

$$Swing(G, i, \chi) \equiv \neg D_G \chi \wedge D_{G \cup \{i\}} \chi$$

- The Banzhaf score of $i$ wrt. goal $\chi$ is at least $k$:

$$BAL(i, k, \chi) \equiv \bigvee_{G_1 \neq \cdots \neq G_k \subseteq Ag \setminus \{i\}} \bigwedge_{G \in \{G_1, \ldots, G_k\}} Swing(G, i, \chi)$$

- The Banzhaf score of $i$ wrt. goal $\chi$ is $k$:

$$B(i, k, \chi) \equiv BAL(i, k, \chi) \wedge \neg BAL(i, k+1, \chi)$$

- Of potential interest is checking whether or not one agent has more information/power than another. Note that the maximal Banzhaf score is determined by the maximum number of coalitions not containing the agent; $2^{n-1}$. The Banzhaf score of agent $i$ is at least as high as that of agent $j$:

$$BNoLower(i, j, \chi) \equiv \bigvee_{k \in [0, 2^{n-1}]} BAL(i, k, \chi) \wedge \neg BAL(j, k, \chi)$$

- $i$ is a veto player wrt. goal $\chi$:

$$Veto(i, \chi) \equiv \neg D_{Ag \setminus \{i\}} \chi$$

  $i$ is a veto player iff it is included in all winning coalitions, iff all coalitions without $i$ are losing, iff $\neg D_G \chi$ holds for all $G$ without $i$. By monotonicity this holds iff $Veto(i, \chi)$ holds.

- $i$ is a dictator wrt. goal $\chi$:

$$Dictator(i, \chi) \equiv Veto(i, \chi) \wedge K_i \chi$$

  $i$ is a dictator iff all coalitions containing $i$ are winning, and no coalition without $i$ is winning. This holds iff $Dictator(i, \chi)$ holds, by monotonicity.

- $i$ is a dummy wrt. goal $\chi$:

$$Dummy(i, \chi) \equiv \bigwedge_{G \in 2^{Ag}} D_{G \cup \{i\}} \chi \rightarrow D_G \chi$$

  $i$ is a dummy iff $\forall G : M, s \models \neg(\neg D_G \chi \wedge D_{G \cup \{i\}} \chi)$ which is equivalent to $\forall G : M, s \models D_{G \cup \{i\}} \chi \rightarrow D_G \chi$.

## 4.2 Full Communication

Implicit in the idea of information-based power is that groups of agents should somehow be able to *realise* the knowledge distributed among them in order to jointly find out that the goal formula is true. However, while distributed knowledge is the most popular concept in the literature aiming to capture the "sum" of the knowledge in a group, it has the following property, as first pointed out in [13]. It might be that $G$ has distributed knowledge of the goal, but it is still not possible for the group to establish $\chi$ through communication in the following sense: it might not be the case that there exists a formula $\varphi_i$ for each $i \in G$ such that $M, s \models \bigwedge_{i \in G} K_i \varphi_i$ and $\models \bigwedge \varphi_i \rightarrow \chi$. This (possibly lacking) communication property is equivalent [13] to:

$$M, s \models D_G \chi \quad \Rightarrow \quad \bigcup_{i \in G} \mathcal{K}_i(M, s) \models \chi \tag{5}$$

and [13] calls this the *principle of full communication* (the other direction of (5), $\bigcup_{i \in G} \mathcal{K}_i(M, s) \models \chi \Rightarrow M, s \models D_G \chi$, holds on any model). As an example, consider the model $M_1$ in Figure 2. In this model $p$ is distributed knowledge among agents 1 and 2 in state $s$, but $p$ is not entailed from the individual knowledge of 1 and 2 in $s$ and the model does not satisfy the principle of full communication.

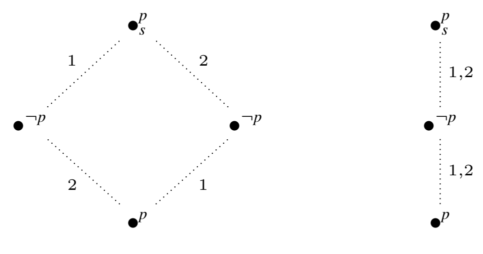

**Figure 2: Models $M_1$ (left) and $M_2$ (right). Reflexive and transitive edges omitted.**

So, if we take the $p$ as the goal formula, agent 1 is swing for $\{2\}$ in state $s$ in the model $M_1$ above, but it is not possible for agents 1 and 2 to actually infer $p$ together by communicating using the epistemic language. Our information-based power measures make particular sense in models that satisfy the principle of full communication, because in such models whatever is distributed knowledge can be obtained by communication in the sense that it follows from individual knowledge that the involved agents can specify and communicate as logical formulas. So which models satisfy the principle of full communication? There are two particularly relevant model properties here (generalisations of propositions given in [13]). A model $M = \langle W, \sim_1, \ldots, \sim_n, \pi \rangle$ is a:

- *full model* [7] iff for all $s \in W$, $G \subseteq Ag$, and $\Phi \subseteq \mathcal{L}_D$: if $\Phi \cup \mathcal{K}_G(M, s)$ is satisfiable then $\Phi$ is satisfiable in $\{t : (s, t) \in \sim_G^D\}$.
- *full communication model* [11] iff for all $s \in W$, $G \subseteq Ag$, and $\varphi \in \mathcal{L}_K$: if $\{\varphi\} \cup \mathcal{K}_G(M, s)$ is satisfiable then $\varphi$ is satisfiable in $\{t : (s, t) \in \sim_G^D\}$.

Clearly, full models are full communication models. [7] shows that fullness is sufficient for the principle of full communication to hold, while [11] shows that a model satisfies the principle of full communication *if and only if* the model is a full communication model.

While this definition of full communication models may seem somewhat technical, note that the principle of full communication is often violated by the existence of bisimilar states in the model (such as in the model above). Indeed, bisimulation contractions of finite models are full communication models (they are *distinguishing* in the sense of [13], due to the existence of characteristic formulae). Models that are finite and do not contain bisimilar states (and thus are their own bisimulation contractions) are very common.

Thus, on full communication models we get an alternative, equivalent, definition of power. We have that:

$$\nu_S^D(G) = 1 \quad \Leftrightarrow \quad \bigcup_{i \in G} \mathcal{K}_i(M, s) \models \chi \tag{6}$$

when $M$ is a full communication model.

## 4.3 Properties of Power

The relationship between power properties and epistemic properties is of natural interest, not the least in order to validate that our definition of power is reasonable. The relationship properties in the following lemma are discussed below.

LEMMA 1. *Let the goal structure $S = \langle M, s, \chi \rangle$ be given.*

1. *If $M, s \models \neg D_{Ag} \chi$, then $x_i = 0$ for all $i$ and $x \in \{\sigma, \mu, \beta, \varsigma\}$.*
2. *If $M, s \models \neg \chi$, then $x_i = 0$ for all $i$ and $x \in \{\sigma, \mu, \beta, \varsigma\}$.*
3. *If $M, s \models K_i \chi$, then $x_i \geq x_j$ for all $j$ and $x \in \{\sigma, \mu, \beta, \varsigma\}$.*
4. *If $M, s \models \neg D_{Ag \setminus \{i\}} \chi$, $x_i \geq x_j$ for all $j$ and $x \in \{\sigma, \mu, \beta, \varsigma\}$.*
5. *If $M, s \models K_i \chi \wedge \neg K_j \chi$, then $x_i > x_j$ for all $x \in \{\sigma, \mu, \beta, \varsigma\}$.*
6. *On full communication models, if $\mathcal{K}_i(M, s) \subseteq \mathcal{K}_j(M, s)$ then $x_i \leq x_j$, for any power measure $x \in \{\sigma, \mu, \beta, \varsigma\}$.*

The first property says that if not enough information to infer the goal formula is distributed throughout the complete system, then every agent has *no power*. The second property is a special case of the first – the goal *cannot* be derived because it is not true. The third and fourth properties represent the other extreme: *maximum power*. The agent has maximum power (at least as much power as anyone else) if she already knows the goal, or if the rest of the system does not have enough information to derive the goal (i.e, if the agent is a veto player). The fifth and sixth properties are about *relative power*. The fifth says that an agent who already knows $\chi$ is always strictly more powerful than an agent who does not know $\chi$. The sixth property says that if one agent knows at least as much as another agent, then the first agent is at least as powerful. This relates our definition of power to a more classical notion of "knowing more" in a reasonable way. Our notion is more fine grained; the implication does not hold in the other direction. The sixth property holds for full communication models, which, again, is a natural class of models in which to interpret our power measures since they come with a natural mechanism for distribution of information.

PROOF (OF LEMMA 1). 1. Follows immediately from monotonicity: if $i$ is swing for $G$, then $M,s \models D_{G\cup\{i\}}\chi$.

2. Immediate from $\models \neg\chi \rightarrow D_{Ag}\neg\chi$ and the first item.

3. If suffices to show that $i$ is swing for any coalition any agent $j$ is swing for. So assume that $M,s \models \neg D_G\chi \wedge D_{G\cup\{j\}}\chi$. From $M,s \models K_i\chi$ it follows that $M,s \models D_{G\cup\{i\}}\chi$, and thus $i$ is also swing for $G$.

4. Assume that $j$ is swing for $G$. From $M,s \models D_{G\cup\{j\}}\chi$, the assumption that $M,s \models \neg D_{Ag\setminus\{i\}}\chi$ and monotonicity, it follows that $i \in G$. Thus it also follows that $i$ is swing for $(G\setminus\{i\})\cup\{j\}$. Because $i \in G$ and $j \notin G$ for coalitions $G$ for which $j$ is swing, $(G_1\setminus\{i\})\cup\{j\} \neq (G_2\setminus\{i\})\cup\{j\}$ for any two different coalitions $G_1, G_2$ for which $j$ is swing, and thus there are at least as many swings for $i$.

5. If $j$ is swing for $G$, $M,s \models \neg D_G\varphi$ so $G$ cannot contain $i$ and $i$ is also swing for $G$. In addition, $i$ is swing for $\emptyset$, unlike $j$.

6. Let $M$ be a full communication model and assume that $i$ is swing for $G$, i.e., that $M,s \models D_G\chi \wedge D_{G\cup\{i\}}\chi$. From the fact that $M$ is a full communication model and eq. (6) above, we get that $\bigcup_{l\in G\cup\{i\}}\mathcal{K}_l(M,s) \models \chi$. From $\mathcal{K}_i(M,s) \subset \mathcal{K}_j(M,s)$ it follows that $\bigcup_{l\in G\cup\{j\}}\mathcal{K}_k(M,s) \models \chi$ which again means that $M,s \models D_{G\cup\{j\}}\chi$. Thus, $j$ is swing for $G$. □

In the following lemma we look at power measures in "similar" models. The proper notion of bisimulation for distributed knowledge, and hence our power measures, is given in the second point.

LEMMA 2.

1. *The power measures are not invariant under (standard) bisimulation. That is, bisimilar pointed models may have different power measures.*
2. *The power measures are invariant under* collective bisimulation *[11].*
3. *On full models, the power measures are invariant under (standard) bisimulation.*

PROOF. 1. A counter-example is found in Figure 2, which contains two bisimilar models with two agents. It is easy to see that by taking $\chi = p$, we get $\sigma_1 = 1$ in $M_1$ but $\sigma_1 = 0$ in $M_2$.

2. follows immediately from the fact that satisfaction in $\mathcal{L}_D$ is invariant under collective bisimulation [11, Prop. 19].

3. For full models the notions of collective bisimulation and bisimulation coincide [11, Prop. 20]. □

Finally, let us look at the relationship between power properties and the structure of the goal formula. We will make use of the logical expressions of power properties from Section 4.1.

Starting with tautologies and contradictions:

$$\begin{array}{ll} \models \neg Swing(G,i,\top) & \models \neg Swing(G,i,\bot) \\ \models Veto(i,\bot) & \models \neg Veto(i,\top) \\ \models \neg Dictator(i,\bot) & \models \neg Dictator(i,\top) \end{array}$$

With such goal formulae, no agent can be swing for any coalition. Every agent is a veto player for $\bot$, while no agent is a veto player for $\top$. No agent can be a dictator for $\bot$ nor $\top$.

The case of conjunction:

$$\models (Swing(G,i,\chi_1) \wedge Swing(G,i,\chi_2)) \rightarrow Swing(G,i,\chi_1 \wedge \chi_2)$$

Swings are closed under the operation of taking conjunction of goal formulae. The converse does not hold, but this does:

$$\models Swing(G,i,\chi_1 \wedge \chi_2) \rightarrow (Swing(G,i,\chi_1) \vee Swing(G,i,\chi_2))$$

– if $i$ is swing wrt. a conjunction, she is swing wrt. at least one of the conjuncts (but not necessarily both).

For negation we have that (but not the other way around):

$$\models Swing(G,i,\neg\chi) \rightarrow \neg Swing(G,i,\chi)$$

Moving on to the case that the goal formula is epistemic, first observe the following properties of distributed $S5$ knowledge: $\models D_G D_{G'}\varphi \rightarrow D_G\varphi$ for any $G, G'$, and $\models D_G D_{G'}\varphi \leftrightarrow D_G\varphi$ when $G \subseteq G'$. From these properties it follows that:

$$\begin{array}{ll} \models Swing(H,i,D_G\chi) \rightarrow Swing(H,i,\chi) & \text{when } H \subseteq G \\ \models Swing(H,i,D_G\chi) \leftrightarrow Swing(H,i,\chi) & \text{when } H\cup\{i\} \subseteq G \end{array}$$

In particular, using a goal formula $D_G\varphi$ is equivalent to using $\varphi$ when it comes to counting swings within $G$.

If we take $G = \{j\}$ in the expressions above, we get the case where the goal formula describes individual knowledge. It follows that:

$$\begin{array}{ll} \models Swing(\emptyset,i,K_j\chi) \rightarrow Swing(\emptyset,i,\chi) & \text{for any } j \\ \models Swing(\{j\},i,K_j\chi) \rightarrow Swing(\{j\},i,\chi) & \text{for any } j \\ \models Swing(\emptyset,i,K_i\chi) \leftrightarrow Swing(\emptyset,i,\chi) & \end{array}$$

# 5. KNOWLEDGE OF POWER

We have thus associated power indices with states of Kripke structures, by assuming that they are defined by agents' knowledge. But epistemic logic allows us to reason about agents' knowledge *about* state-properties – so we can go from analysing the power of knowledge to analysing knowledge of power: what do the agents in the system know about the distribution of power?

The formula $K_jSwing(G,i,\chi)$, where $Swing(G,i,\chi) = \neg D_G\chi \wedge D_{G\cup\{i\}}\chi$, denotes the fact that agent $j$ knows that $i$ is swing for $G$. If we look first at the more general case of *distributed* knowledge of that fact, we have the following (we formally prove this and the following validities in Theorem 2 below):

$$\models Swing(G,i,\chi) \rightarrow D_{G\cup\{i\}}Swing(G,i,\chi) \quad (7)$$

– if $i$ is swing for $G$, then this is distributed knowledge in $G\cup\{i\}$.

However, this does not carry over to individual knowledge. It turns out that $Swing(G,i,\chi) \wedge \neg K_jSwing(G,i,\chi)$ is satisfiable, for any $j$ including $j = i$. Thus, an agent can be swing for a coalition, without neither the agent nor the agents in the coalition knowing it. When, then, *does* an agent know that she is swing? The answer is: *almost never*. The following holds:

$$\models K_j\neg Dummy(i,\chi) \rightarrow K_j\chi \quad (8)$$

for any $i,j$ (including $i = j$). In other words, an agent can only know that any agent (including herself) is swing for any coalition if she (the first agent) already knows the goal formula! In the typical case that $\chi$ is distributed information throughout the system, but no individual agent alone knows $\chi$, *no* agent knows that *any* agent can swing *any* coalition from ignorance to knowledge about $\chi$. It follows that

$$\models K_j\neg Dummy(i,\chi) \rightarrow K_j \bigwedge_{k\in Ag} BNoLower(j,k,\chi) \quad (9)$$

– only agents that are maximally powerful (at least as powerful as any other agent), and know that they are, can know that anyone (including themselves) are not a dummy player.

It also holds that

$$\models K_j Swing(G,i,\chi) \rightarrow K_j Swing(G,j,\chi) \quad (10)$$

– if an agent knows that another agent is swing for some coalition, then the first agent must be swing for the same coalition. In particular: $\models K_j \neg Dummy(i,\chi) \rightarrow K_j \neg Dummy(j,\chi)$.

However, *no* agent *in* a coalition can know that someone is swing for that coalition:

$$\models \bigwedge_{j \in G} \neg K_j Swing(G,i,\chi) \quad (11)$$

For veto players, we have that

$$\models K_i Veto(j,\chi) \rightarrow \neg K_i \neg Dummy(i,\chi) \quad i \neq j \quad (12)$$

– the only agents that can know that someone else is a veto player are agents that consider it possible that they are dummies themselves.

For dictators, we have that

$$\models \neg K_j Dictator(i,\chi) \quad i \neq j \quad (13)$$

– the only agent that can know who the dictator is, is the dictator.

Turning to knowledge about the values of power indices, we have

$$\models K_j B(i,k,\chi) \rightarrow BNoLower(j,i,\chi) \quad (14)$$

– no agent can know the Banzhaf score of any agent with a lower score than herself.

We can conclude that the distribution of power is generally not known *in* the system. We emphasise that this does not pose any problem for our interpretation of the power indices as measures of the distribution of information in the system, as we discuss further in Section 7.

THEOREM 2. *Properties (7)–(14) hold.*

PROOF. We make use of the fact that distributed knowledge satisfies the $S5$ properties [4], which follows from the fact that the intersection of equivalence relations is an equivalence relation, as well as the monotonicity property ($D_G\varphi \rightarrow D_H\varphi$ when $G \subseteq H$).

(7): from $\neg D_G\chi$ it follows that $D_G\neg D_G\chi$ by negative introspection, and $D_{G\cup\{i\}}\neg D_G\chi$ follows by monotonicity. $D_{G\cup\{i\}}D_{G\cup\{i\}}\chi$ follows from $D_{G\cup\{i\}}\chi$ by positive introspection. $D_{G\cup\{i\}}Swing(G,i)$ follows by knowledge distribution.

(8): $K_j\neg Dummy(i,\chi)$ is equal to $K_j \bigvee_G (D_{G\cup\{i\}}\chi \wedge \neg D_G\chi)$. By reflexivity $D_{G\cup\{i\}}\chi$ implies $\chi$, and thus $\bigvee_G (D_{G\cup\{i\}}\chi \wedge \neg D_G\chi)$ implies that $\chi$. By knowledge distribution, $K_j\chi$ holds.

(9): let $K_j\neg Dummy(i,\chi)$ be true. By (8), $K_j\chi$ and from positive introspection $K_jK_j\chi$. From Lemma 1.3 it follows that $K_jBNoLower(j,k,\chi)$ for any $k$.

(10): from $K_jSwing(G,i,\chi)$ it follows that $K_j\neg D_G\chi$. By (8) it also follows that $K_j\chi$. By knowledge distribution, $K_j(\neg D_G\chi \wedge K_j\chi)$, which by monotonicity implies that $K_j(\neg D_G\chi \wedge D_{G\cup\{j\}}\chi)$.

(11): if $K_jSwing(G,i,\chi)$ is true for some $j \in G$, then $K_jSwing(G,j,\chi)$ by (10), and $Swing(G,j,\chi)$ by reflexivity. But this is a contradiction.

(12): from $K_iVeto(j,\chi)$ it follows that $K_i\neg K_i\chi$ when $i \neq j$, from which it follows that $\neg K_i\chi$. If $K_i\neg Dummy(i,\chi)$ is true, then $K_i\chi$ by (8); a contradiction.

(13): $K_jDictator(i,\chi)$ is equivalent to $K_j(Veto(i,\chi) \wedge K_i\chi)$, which implies that $K_j\chi$ and $Veto(i,\chi)$. From the latter it follows that $\neg D_{Ag\setminus\{i\}}\chi$, and from monotonicity it follows that $\neg K_j\chi$ – a contradiction.

(14): if $\sigma_i = 0$, the formula holds trivially. If $\sigma_i > 0$, $K_jB(i,k,\chi)$ implies that there is a $G$ such that $K_j(\neg D_G\chi \wedge D_{G\cup\{i\}}\chi)$ is true. It follows that $K_j\chi$, and by Lemma 1.3 that $\sigma_j \geq \sigma_i$. □

# 6. OTHER TYPES OF GROUP KNOWLEDGE

We have so far used the notion of distributed knowledge to measure power. Can other notions of group knowledge be used? Note that both everybody-knows and common knowledge are anti-monotonic, in the sense that $C_G\varphi$ implies $C_{G'}\varphi$ when $G' \subseteq G$, while distributed knowledge is monotonic ($D_{G'}\varphi$ implies $D_G\varphi$). This means that simply "replacing" distributed knowledge in the definition of the game by any of these notions would not make sense (e.g., $\neg C_G\varphi \wedge C_{G\cup\{i\}}\varphi$ is not satisfiable). However, there is another way in which we can look at an agent's power with respect to common knowledge (and similarly with everybody-knows). An agent has "negative" power if he can swing a coalition from *having* common knowledge of the goal, to *not* having it. In other words, this would correspond to an agent's power to spoil, rather than to achieve, the goal. Using this definition of the power measures, a high value means that the agent has *little* information, and including it in a group is likely to, e.g., break common knowledge needed for coordination.

Let us start with everybody-knows. Given $S = \langle M,s,\chi\rangle$, let:

$$\nu_S^E(G) = \begin{cases} 1 & M,s \models \neg E_G\chi \\ 0 & \text{otherwise} \end{cases}$$

We say that a simple cooperative game is *determined* if there is a set of agents $Winners \subseteq Ag$ such that $\nu(G) = 1$ iff $G \cap Winners \neq \emptyset$. Note that determined games are monotonic.

THEOREM 3. *For any simple cooperative game $\Gamma = \langle Ag, \nu\rangle$, there exists a goal structure $S$ such that $\nu_S^E = \nu$ iff $\Gamma$ is determined.*

PROOF. For the implication to the right, given $S$ let $Winners = \{i : M,s \models \neg K_i\chi\}$. It is easy to see that $\nu_S^E(G) = 1$ iff $G \cap Winners \neq \emptyset$. For the implication to the right, we define $S = \langle M,s,\chi\rangle$ as follows. Let $p \in \Theta$. Let $W = \{s,t\}$; $s_0 = s$; $V(p) = \{s\}$, $V(q) = \emptyset$ for $q \neq p$; $s \sim_i t \Leftrightarrow i \in Winners$; $\chi = p$. Let $\nu(G) = 1$. That means that there is an agent $i$ such that $i \in G \cap Winners$. From $i \in Winners$ it follows that $M,s_0 \models \neg K_ip$, and since $i \in G$ we get that $M,s_0 \models \neg E_G\chi$. For the other direction, let $M,s_0 \models \neg E_Gp$. That means that $M,s_0 \models \neg K_ip$ for some $i \in G$. But the only possibility then is that also $i \in Winners$. Thus, $i \in G \cap Winners$, and thus $\nu(G) = 1$. □

It is easy to see that for determined games, the Banzhaf score is the same for all winners, as well as the same (0) for all non-winners:

LEMMA 3. *For any determined game and any agent i,*

$$\sigma_i = \begin{cases} 2^{|Ag\setminus Winners|} & i \in Winners \\ 0 & otherwise \end{cases}$$

It follows that it is easy to compute the power measures:

THEOREM 4. *Given a goal structure $S = \langle M,s,\chi\rangle$ and an agent i in M, the Banzhaf score $\sigma_i$ for i in the game $\langle Ag, \nu_S^E\rangle$ can be computed in polynomial time.*

PROOF. By Theorem 3 the game is determined. The winners can be computed in polynomial time: for every state $t$, check whether $M,t \models \neg\chi$, and if it does add $i$ to $Winners$ if there is an $i$-transition from $t$ to $s$. $\sigma_i$ is computed from the size of $Winners$ according to Lemma 3. □

Moving on to common knowledge, given $S = \langle M, s, \chi \rangle$, let:

$$\nu_S^C(G) = \begin{cases} 1 & M, s \models \neg C_G \chi \\ 0 & \text{otherwise} \end{cases}$$

EXAMPLE 2. *The following two examples are inspired by [14, Section 2.3]. In the first setting, the set of agents Ag is the set of participants of a conference, and $a \in Ag$ represents our hero Alco. During one afternoon, while all other participants are attending a joint session, Alco spends his time in the bar of the conference hotel. The session chair announces $\chi$: 'tomorrow, sessions start at 9:00 rather than 10:00'. Everybody (i.e., Ag) at the conference feels very responsible for the well-being of the participants, and only if $C_{Ag}\chi$ holds, people will stop informing each other of $\chi$. If s is the situation immediately after the chair's announcement, we obviously have $M, s \models Swing(Ag \setminus \{a\}, a, \chi)$, where Swing is now defined for common knowledge: $Swing(G, i, \chi) = C_G\chi \wedge \neg C_{G\cup\{i\}}\chi$. Now consider a new state $s_1$, in which Alco leaves the bar to get some fresh air, and which leads to a state $s_2$ where at the general session a friend f of Alco makes the chair (publicly) aware that Alco was in the bar during the announcement $\chi$. At this moment it is common knowledge among $Ag \setminus \{a\}$ that $Swing(Ag \setminus \{a\}, a, \chi)$, but then the chair replies to f by saying that there is an intercom in the bar that is directly connected to the conference room. Note that a is now still a veto player wrt. Ag and $\chi$, since Alco does not know about the discussion regarding his absence during the announcement of $\chi$. In other words, although in $s_2$ we have $E_{Ag}\chi$, we also have $\neg K_a K_f K_a \chi$: Alco knows that his friend f may have concerns about Alco not knowing $\chi$ (this concern is justified, since f notified the chair), and Alco does not know that f has been properly informed (that $K_a\chi$) by the chair, so one may expect that a will make at some time an effort to make publicly known that he knows $\chi$, so people can stop worrying about a's time-table tomorrow.*

*Swing players for common knowledge in a coalition G often come with delicate protocols for the communication in G. An example here is the celebrations of Santa Claus in certain cultures, where it is common knowledge among those over a certain age that Santa Claus is in fact not responsible for the presents at the evening (this is $\chi$), while $\chi$ is not known among the participants under a certain age. Now, even when everybody at the Christmas party knows that $\chi$, there may be several swing players for several coalitions, which explains that conversations have to be participated in carefully. To be more precise, suppose that $E_G E_G \chi \wedge \neg K_i K_j K_i \chi$ (with $i, j \in G$). Since i knows that everybody in G knows $\chi$ already, he might chose not to look childish to j and reveal to j that $K_i\chi$, indicating he is not a fool. But i might also chose to exploit $\neg K_i K_j K_i \chi$, and challenge j into a 'dangerous conversation', where j may think he needs to be careful not to reveal $\chi$ to i.*

*These examples also suggest that power is in fact an interesting issue in dynamic contexts, after enough communication has taken place for instance, Alco may seize to be a swing player. Dynamic Epistemic Logic ([14]) paves the right formal framework to study these phenomena, like the fact that some true formulas can never be known no matter how often they are announced: they would always have a veto player (Moore sentences like $(p \wedge \neg K_a p)$ being the most prominent examples).*

Like for the case of distributed knowledge, the class of games obtained in this way is exactly the monotonic games.

THEOREM 5. *For any simple cooperative game $\Gamma = \langle Ag, \nu \rangle$, there exists a goal structure S such that $\nu_S^C = \nu$ iff $\Gamma$ is monotonic.*

PROOF. It is easy to see that $\nu_S^C$ is monotonic.

For the other direction, let $\nu$ be monotonic. If there is no coalition $G$ with $\nu(G) = 1$, let $M$ consist of only one state $s$ with $V(p) = \{s\}$ and $\sim_a = W \times W$ for every $a \in Ag$. It is easily seen that $\nu_{M,s,p}^C(G) = 0$ for all coalitions $G$.

Otherwise put first of all $s \in W \cap V(p)$ and add $(s, s)$ to each $\sim_a$. Let $H_1, \ldots H_k$ be the coalitions with the property that $\nu(H_i) = 1$ and for no proper subset of $H_i$, it holds that $\nu(H) = 1$. For each such $H_i$, do the following. Let $H_i = \{a_1^i, a_2^i, \ldots a_{m(i)}^i\}$. Add new states $W_i = \{s_1^i, s_2^i, \ldots s_{m(i)}^i\}$ to $W$ in such a way that $(s, s_1^i)$ and $(s_1^i, s)$ become members of $\sim_{a_1^i}$ and furthermore add $(s_j^i, s_{j+1}^i)$, $(s_{j+1}^i, s_j^i)$ to $\sim_{a_{j+1}^i}$ with $1 \le j < m(i)$. Add $(s_j^i, s_j^i)$ to each $\sim_a$ $(1 \le m(i))$. Finally, add $W_i \setminus \{s_{m(i)}^i\}$ to $V(p)$. When this process has finished for all $H_i$, take the transitive symmetric reflexive closure of every $\sim_a$ so far defined. The effect of this last step is that for every agent $a$ and every two states $s_1^i$ and $s_1^j$ with $(s, s_1^i)$ and $(s, s_1^j) \in \sim_a$, we also add $(s_1^i, s_1^j)$ and $(s_1^j, s_1^i)$ to $\sim_a$.

A straight path $\pi$ in the model is a sequence of state-agent alterations $\langle x_1, a_1, x_2, a_2 \ldots x_n \rangle$, with each $x_i \in W, a_i \in Ag$, and $(x_i, x_{i+1}) \in \sim_{a_i}$ such that $x_i \ne x_j$ if $i \ne j$. It is a straight $s$-path if $x_1 = s$. Let $Ag(\pi)$ be the set of agents occurring in $\pi$. Note that a straight $s$-path that ends in state $s_n$ denotes a 'shortest' route in the model from $s$ to $s_n$, since the states in a straight path are different. A straight path $x_1, a_1, x_2, a_2 \ldots x_n$ leads to $\varphi$ if $x_n$ is the only-$\varphi$ world in it. The following is an important property of our model: there is a straight path $\pi$ leading to $\neg p$ iff for some $H_i$, we have $\nu(H_i) = 1$ and $Ag(\pi) = H_i$.

We now prove that $\forall G \subseteq Ag$ ($\nu(G) = 1$ iff $M, s \models \neg C_G p$). First, if $\nu(G) = 1$, there is a smallest set $H_i = \{a_1^i, \ldots, a_{m(i)}^i\} \subseteq G$ such that $\nu(H_i) = 1$. For this $H_i$, we have constructed a straight $s$-path $\pi$ leading to $\neg p$ and for which $Ag(\pi) = H_i$. So, we have $M, s \models \neg C_{H_i} p$, and hence $M, s \models \neg C_G p$, i.e., $\nu_S^C(G) = 1$. Secondly, suppose $M, s \models \neg C_G p$, it means for our model that there is a straight $s$-path $\pi$ leading to $\neg p$ for which $Ag(\pi) \subseteq G$ (indeed, there may be agents $a \in G \setminus Ag(\pi)$). But the only such paths we have in $M$ are paths that use a minimal set of agents $H_i$ for which $\nu(H_i) = 1$, so $\nu(Ag(\pi)) = 1$. By monotonicity, $\nu(G) = 1$. □

# 7. DISCUSSION

We have shown that our information-based notion of power has reasonable properties, at least on full communication models – which come with a natural mechanism for distribution of information. We have also shown that it is easy to compute such power indices using a standard model checker for epistemic logic.

It is natural to define swings using distributed knowledge. A high power index here means that the agent's knowledge is important for an arbitrary group jointly getting to know the goal formula by sharing their information. We also gave alternative definitions of "negative" power in terms of swinging a group from a situation where every member knows the goal, or the goal is common knowledge. Here, a high power index means that the agent knows little: if it is important to have common knowledge in a group (e.g., for coordination), then it is likely that including a high-power agent will lead to failure. The everybody-knows case is computationally tractable, but the price is a lower "resolution": the agents divide into only two classes, with agents in the same class having the same power. It is interesting that the common knowledge case and the distributed knowledge correspond to the same class of voting games (Theorems 1 and 5). If this seems counter-intuitive, keep in mind that the two theorems express that there is a connection between distributed knowledge and *lack* of common knowledge: conceiving distributed knowledge as a game where a coalition wins if it implicitly knows the goal formula, is structurally similar to conceiving common knowledge as a game where a coalition wins if it does *not* commonly know the goal.

[15] studies a particular notion of "knowing more". Their concept "$i$ knows at least what $j$ knows" is defined by $R_i(s) \subseteq R_j(s)$ where $s$ is a state and $R_x(s) = \{t : (s,t) \in R_x\}$ and $R_x$ is an indistinguishability relation for agent $x$. Our power measures for distributed knowledge agree: if $R_i(s) \subseteq R_j(s)$ then $Swing(G,j,\chi)$ implies that $Swing(G,i,\chi)$ for any $\chi$, and thus $\sigma_i \geq \sigma_j$. The implication does not hold in the other direction; our notion of "knowing more" is more fine grained. [15] also introduces a modal operator $\succeq$ where, for agents $i$ and $j$, the formula $i \succeq j$ expresses that whatever state is an alternative for $j$, is also an alternative for $i$. This provides a way to locally express that $K_i\varphi \rightarrow K_j\varphi$ for all $\varphi$. There is one sense in which such an operator allows one also to express properties of the power of knowledge in a compact way. For distributed knowledge for instance, the formula $i \succeq j$ implies that $(Swing(G,i,\chi) \rightarrow Swing(G,j,\chi))$ and $\neg Swing(G \cup \{i\},j,\chi)$ – for *any* $\chi$. When reasoning about the power in the context of everybody knows, "opposite" properties derive: $\models (i \succeq j) \rightarrow (Swing(G,j,\chi) \rightarrow Swing(G,i,\chi))$ and $\models (i \succeq j) \rightarrow \neg Swing(G \cup \{j\},i,\chi)$. Note that such properties cannot be expressed in modal logic without such an operator: for instance in $\models (K_i\varphi \rightarrow K_j\varphi) \rightarrow (Swing(G,i,\chi) \rightarrow Swing(G,j,\chi))$ the formula $\varphi$ is a specific formula (not a scheme), and $\models (K_i\varphi \rightarrow K_j\varphi) \Rightarrow \models (Swing(G,i,\chi) \rightarrow Swing(G,j,\chi))$ is obviously true, but much weaker: the antecedent is false (if $i \neq j$).

In Section 5 we saw that agents *in* the system generally know very little about the distribution of information-based power in the system. For example, an agent with a high power index typically does not know which coalitions she needs to join in order to derive the goal formula (or indeed *that* she is a high-power agent). We emphasise that this is not in any way a problem for the interpretation of our power indices. A high Banzhaf index means, in our setting, that the probability of changing some arbitrary coalition from ignorance to knowledge about the goal is high – in the same way that it is interpreted as the probability of changing an outcome in voting theory. In fact, that an agent does not know which coalitions it is swing for makes the probability of being swing for an *arbitrary* coalition more interesting. Furthermore, in many distributed and multi-agent systems, such as sensor networks, agents are restricted to communication with some arbitrary sub-group of all agents at any given time. We think of these power measures as a tool for external analysis of the information distribution in a system, to find out, e.g., whether information is evenly distributed or whether there are some agents that are particularly crucial to the functioning of the system in the sense that the information they have is difficult to obtain elsewhere in the system. The negative results about knowledge of power properties can also be seen as a *barrier against strategic behaviour*: it is almost never possible for an agent to know that it suffices to share information with only some particular subgroup of the grand coalition.

An interesting direction for future work is to associate formulae of the form $D_G D_H \varphi$ with *composite* voting games [6, p. 27]. In this paper we have studied a semantic notion of power, associated with a point in a Kripke structure. Another direction for future work is to develop a *syntactic* notion of power, based on a set of epistemic formulae. For such an approach it would be necessary to syntactically describe that agents know "this and nothing more", and extensions of epistemic logic with *only knowing* [9] seem like a promising starting point.

## Acknowledgments

We thank the AAMAS program committee and Pål Grønås Drange for comments that helped us improve the paper.

## 8. REFERENCES

[1] T. Ågotnes, W. van der Hoek, J. A. Rodriguez-Aguilar, C. Sierra, and M. Wooldridge. On the logic of normative systems. In M. M. Veloso, editor, *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI 2007)*, pages 1175–1180, California, 2007. AAAI Press.

[2] T. Ågotnes, W. van der Hoek, M. Tennenholtz, and M. Wooldridge. Power in normative systems. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 145–152, Budapest, Hungary, May 2009. IFAMAAS.

[3] J. F. Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19(2):317–343, 1965.

[4] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press: Cambridge, England, 2001.

[5] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning About Knowledge*. The MIT Press: Cambridge, MA, 1995.

[6] D. S. Felsenthal and M. Machover. *The Measurement of Voting Power*. Edward Elgar: Cheltenham, UK, 1998.

[7] J. Gerbrandy. *Bisimulations on Planet Kripke*. Ph.D. thesis, University of Amsterdam, 1999.

[8] A. Hunter and S. Konieczny. On the measure of conflicts: Shapley inconsistency values. *Artificial Intelligence*, 174(14):1007 – 1026, 2010.

[9] H. J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence*, 42:263–309, 1990.

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

[11] F. Roelofsen. Distributed knowledge. *Journal of Applied Non-Classical Logics*, 17(2):255–273, 2007.

[12] Y. Shoham and M. Tennenholtz. On the synthesis of useful social laws for artificial agent societies. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, San Diego, CA, 1992.

[13] W. van der Hoek, B. van Linder, and J.-J. Meyer. Group knowledge is not always distributed (neither is it always implicit). *Mathematical Social Sciences*, 38:215–240, 1999.

[14] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Springer, Berlin, 2007.

[15] H. Van Ditmarsch, W. Van Der Hoek, and B. Kooi. Knowing more: from global to local correspondence. In *IJCAI'09: Proceedings of the 21st international jont conference on Artifical intelligence*, pages 955–960, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.