Strategic Coalitions with Perfect Recall

Pavel Naumov
pnaumov@vassar.edu

Jia Tao
Lafayette College, taoj@lafayette.edu

Follow this and additional works at: https://digitalwindow.vassar.edu/faculty_research_reports
Part of the Artificial Intelligence and Robotics Commons, Epistemology Commons, Logic and Foundations Commons, Logic and Foundations of Mathematics Commons, and the Theory and Algorithms Commons

Citation Information
Naumov, Pavel and Tao, Jia, "Strategic Coalitions with Perfect Recall" (2018). Faculty Research and Reports. 104.
https://digitalwindow.vassar.edu/faculty_research_reports/104
Strategic Coalitions with Perfect Recall

Pavel Naumov  
Department of Computer Science  
Vassar College  
Poughkeepsie, New York 12604  
pnaumov@vassar.edu

Jia Tao  
Department of Computer Science  
Lafayette College  
Easton, Pennsylvania 18042  
taoj@lafayette.edu

Abstract
The paper proposes a bimodal logic that describes an interplay between distributed knowledge modality and coalition know-how modality. Unlike other similar systems, the one proposed here assumes perfect recall by all agents. Perfect recall is captured in the system by a single axiom. The main technical results are the soundness and the completeness theorems for the proposed logical system.

Introduction
Autonomous agents such as self-driving cars and robotic vacuum cleaners are facing the challenge of navigating without having complete information about the current situation. Such a setting could be formally captured by an epistemic transition system where an agent uses instructions to transition the system between states without being able to distinguish some of these states. In this paper we study properties of strategies in such systems. An example of such a system is the epistemic transition system $T_1$, depicted in Figure 1. It has six states named $w_0, w'_0, w_1, w'_1, w_2, w'_2$ and two instructions 0 and 1 that an agent $a$ can use to transition the system from one state to another. For instance, if an instruction 0 is given in state $w_0$, then the system transitions into state $w_1$. The system is called epistemic because the agent cannot distinguish state $w_i$ from state $w'_i$ for each $i = 0, 1, 2$. The indistinguishability relation is shown in the figure using dashed lines. Atomic proposition $p$ is true only in state $w_2$.

![Diagram](image)

Figure 1: Epistemic Transition System $T_1$.

The logical system that we propose consists of two modalities. The first is the knowledge modality $K$. Imagine that the system starts in state $w_2$. Since agent $a$ cannot distinguish state $w_2$ from state $w'_2$ where statement $p$ is not satisfied, the agent does not know if $p$ is true or not. We write this as $(w_2) \not K_a p$. Next, suppose that the system started in state $w_1$ and the agent used instruction 0 to transition the system into state $w_2$. In this paper we assume that all agents have perfect recall, so in state $w_2$ the agent remembers history $(w_1, 0, w_2)$. However, such a history is indistinguishable from history $(w'_1, 0, w'_2)$ because the agent cannot distinguish state $w_1$ from state $w'_1$ and state $w_2$ from state $w'_2$. Thus, the agent does not know that proposition $p$ is true in the state $w_2$ even with history $(w_1, 0, w_2)$. We denote this by $(w_1, 0, w_2) \not K_a p$. Finally, assume that the system started in state $w_0$ and the agent first used instruction 1 to transition it into state $w_1$ and later instruction 0 to transition it to state $w_2$. Thus, the history of the system is $(w_0, 1, w_1, 0, w_2)$. The only history that the agent cannot distinguish from this one is history $(w'_1, 0, 1, 0, w_2)$. Since both of these histories end in a state where proposition $p$ is satisfied, agent $a$ does know that proposition $p$ is true in state $w_2$, given history $(w_0, 1, w_1, 0, w_2)$. We write this as $(w_0, 1, w_1, 0, w_2) \models K_a p$.

The other modality that we consider is the strategic power. In system $T_1$, the agent can transition the system from state $w_1$ to state $w_2$ by using instruction 0. Similarly, the agent can transition the system from state $w'_1$ to state $w'_2$ by using instruction 1. In other words, given either history $(w_1)$ or history $(w'_1)$ the agent can transition the system to a state in which atomic proposition $p$ is satisfied. We say that, given either history, agent $a$ has a strategy to achieve $p$. Histories $(w_1)$ and $(w'_1)$ are the only histories indistinguishable by agent $a$ from history $(w_1)$. Since she has a strategy to achieve $p$ under all histories indistinguishable from history $(w_1)$, we say that given history $(w_1)$ the agent knows that she has a strategy. Similarly, given history $(w'_1)$, she also knows that she has a strategy. However, since indistinguishable histories $(w_1)$ and $(w'_1)$ require different strategies to achieve $p$, given history $(w_1)$ she does not know what the strategy is. We say that she does not have a know-how strategy. We denote this by $(w_1) \not H_a p$, where $H$ stands for know-How. Of course, it is also true that $(w'_1) \not H_a p$.

The situation changes if the transition system starts in state $w_0$ instead of state $w_1$ and transitions to state $w_1$ under instruction 1. Now the history is $(w_0, 1, w_1)$ and the histo-
ries that the agent cannot distinguish from this one are history \((w'0, 1, w_3)\) and history \((w_0, 1, w_1)\) itself. Given both of these two histories, agent \(a\) can achieve \(p\) using the same transition 0. Thus, \((w_0, 1, w_1) \models H_ap\).

Finally note that there are only two histories: \((w_0)\) and \((w'_0)\) indistinguishable from \((w_0)\). Given either history, agent \(a\) can achieve \(H_ap\) using instruction 1. Thus, \((w_0) \models H_aH_ap\). That is, given history \((w_0)\) agent \(a\) knows how to transition to a state in which formula \(H_ap\) is satisfied.

**Multiagent Systems** Like many other autonomous agents, self-driving cars are expected to use vehicle-to-vehicle communication to share traffic information and to coordinate actions (Harding et al. 2014). Thus, it is natural to consider epistemic transition systems that have more than one agent. An example of such a system \(T_2\) is depicted in Figure 2. This system has five epistemic states: \(w_0, w_1, w_2, w_3,\) and \(w_4\) and three agents: \(a, b,\) and \(c.\) In each state the agents vote either 0 or 1 and the system transitions into the next state based on the majority vote. For example, since the directed edge from state \(w_0\) to state \(w_2\) is labelled with 1, if the majority of agents in state \(w_0\) votes 1, then the system transitions into state \(w_2\). Since coalition \(\{a, b\}\) forms a majority, this coalition has a strategy to transition the system from \(w_0\) to state \(w_4\) and, thus, to achieve \(p\). Note that agent \(a\) cannot distinguish state \(w_0\) from state \(w_1\) and thus agent \(a\) does not know what she should vote for to achieve \(p\) because she cannot distinguish state \(w_0\) from state \(w_2\). In this paper, we assume that members of a coalition make the decisions based on combined (distributed) knowledge of the whole coalition. In our example, coalition \(\{a, b\}\) can distinguish state \(w_0\) from both state \(w_1\) and state \(w_2.\) Thus, given history \((w_0)\) the coalition \(\{a, b\}\) knows how to achieve \(p\). We denote this by \((w_0) \models H_{\{a,b\}}p\), or simply as \((w_0) \models H_{a,b}p\).

**Universal Principles** We have discussed a statement being true or false given a certain history. This paper focuses on the logical principles that are true for each history in each epistemic transition system. An example of such a principle is the *strategic positive introspection*: \(H_C\varphi \rightarrow KC_HC\varphi\). This principle says that if a coalition knows how to achieve \(\varphi\), then the coalition knows that it knows how to achieve \(\varphi\).

Informally, this principle is true because in order for statement \(H_C\varphi\) to be satisfied for a given history \(h\), coalition \(C\) must have a strategy to achieve \(\varphi\) that works under any history \(h'\) indistinguishable from history \(h\) by the coalition. Thus, the same strategy must work for any history \(h''\) indistinguishable from history \(h'\) by the coalition. In other words, it is also true that \(h' \models HC\varphi\). Recall that \(h'\) is an arbitrary history indistinguishable from history \(h\) by coalition \(C\). Hence, \(h \models KC\varphi\). According to the standard semantics of the epistemic modality \(KC\), a similar argument can be used to justify the *strategic negative introspection*: \(¬HC\varphi \rightarrow KC¬HC\varphi\).

Another universal principle is the *empty coalition principle*: \(K_{\emptyset} \varphi \rightarrow H_{\emptyset} \varphi\). Indeed, \(K_{\emptyset} \varphi\) means that statement \(\varphi\) is true under any history indistinguishable from the given history by an empty coalition. Since an empty coalition cannot distinguish any two histories, the assumption \(K_{\emptyset} \varphi\) means that statement \(\varphi\) is true under any history. In particular, this statement is true after the next transition no matter how agents vote. Hence, \(H_{\emptyset} \varphi\).

The epistemic modality \(KC\) also satisfies axioms of epistemic logic S5 for distributed knowledge. Know-how modality satisfies the *unachievability of falsehood* principle: \(¬HC⊥\), stating that no coalition can achieve \(⊥\). Know-how modality also satisfies a form of cooperation principle (Pauly 2001; 2002):

\[
H_C(\varphi \rightarrow \psi) \rightarrow (H_D \varphi \rightarrow H_{C\cup D} \psi), \text{ where } C \cap D = \emptyset.
\]

**Perfect Recall** A complete trimodal logical system describing the interplay between distributed knowledge modality \(KC\), coalition know-how modality \(HC\), and standard (not know-how) strategic power modality in the *imperfect recall* setting was proposed by (Naumov and Tao 2017c). We provide a complete axiomatization of the interplay between modalities \(KC\) and \(HC\) in the *perfect recall* setting. Surprisingly, the assumption of perfect recall by all agents is captured by a single principle that we call the *perfect recall principle*: \(H_D \varphi \rightarrow H_D KC\varphi\), where \(D \subseteq C \neq \emptyset\). This principle says that if a sub-coalition \(D \subseteq C\) can achieve \(\varphi\), then after the vote the whole coalition will know that \(\varphi\) is true. Informally, this principle is true because coalition \(C\) is able to recall how sub-coalition \(D\) voted and, thus, will deduce that formula \(\varphi\) is true after the transition. As an empty coalition has no memory even in the perfect recall setting, it is essential for coalition \(C\) to be nonempty. However, the sub-coalition \(D\) can be empty.

**Literature Review** Non-epistemic logics of coalition power were developed by (Pauly 2001; 2002), who also proved the completeness of the basic logic of coalition power. His approach has been widely studied in the literature (Goranko 2001; van der Hoek and Wooldridge 2005; Bogo 2007; Sauro et al. 2006; Ågotnes et al. 2010; Ågotnes, van der Hoek, and Wooldridge 2009; Belardinelli 2014; Goranko, Jamroga, and Turrini 2013). An alternative logical system for coalition power was proposed by (More and Naumov 2012).

(Alur, Henzinger, and Kupferman 2002) introduced Alternating-Time Temporal Logic (ATL) that combines
temporal and coalition modalities. (van der Hoek and Wooldridge 2003) proposed to combine ATL with epistemic modality to form Alternating-Time Temporal Epistemic Logic. They did not prove the completeness theorem for the proposed logical system. Aminof, Murano, Rubin and Zuleger (Aminof et al. 2016) studied model-checking problems of an extension of ATL with epistemic and “prompt eventually” modal operators.

(Ågotnes and Alechina 2012) proposed a complete logical system that combines the coalition power and epistemic modalities. Since their system does not have epistemic requirements on strategies, it does not contain any axioms describing the interplay of these modalities. In the extended version, (Ågotnes and Alechina 2016) added a complete axiomatization of an interplay between single-agent knowledge and know-how modalities.

Know-how strategies were studied before under different names. While (Jamroga and Ågotnes 2007) talked about “knowledge to identify and execute a strategy”, (Jamroga and van der Hoek 2004) discussed “difference between an agent knowing that he has a suitable strategy and knowing the strategy itself”. (van Bentham 2001) called such strategies “uniform”. (Broersen 2008) investigated a related notion of “knowingly doing”, while (Broersen, Herzig, and Troquard 2009) studied modality “know they can do”. (Wang 2015; 2016) captured the “knowing how” as a binary modality in a complete logical system with a single agent and without the knowledge modality.

Coalition know-how strategies for enforcing a condition indefinitely were investigated by (Naumov and Tao 2017a). Such strategies are similar to (Pauly 2001, p. 80) “goal maintenance” strategies in “extended coalition logic”. A similar complete logical system in a single-agent setting for know-how strategies to achieve a goal in multiple steps rather than to maintain a goal is developed by (Fervari et al. 2017).

(Naumov and Tao 2017c) also proposed a complete trimodal logical system describing an interplay between distributed knowledge modality $K^c$, coalition know-how modality $H^c$, and standard (not know-how) strategic power modality in the imperfect recall setting.

In this paper we provide a complete axiomatization of an interplay between modalities $K^c$ and $H^c$ in the perfect recall setting. The main challenge in proving the completeness, compared to (Ågotnes and Alechina 2016; Fervari et al. 2017; Naumov and Tao 2017c; 2017a), is the need to construct not only “possible worlds”, but the entire “possible histories”, see the proof of Lemma 15.

**Outline** First, we introduce the syntax and semantics of our logical system. Next, we list the axioms and give examples of proofs in the system. Then, we prove the completeness of this system. The proof of the soundness is available in the full version of this paper (Naumov and Tao 2017b).

**Syntax and Semantics** Throughout the rest of the paper we assume a fixed set of agents $\mathcal{A}$. By $X^Y$ we denote the set of all functions from set $Y$ to set $X$, or in other words, the set of all tuples of elements from set $X$ indexed by the elements of set $Y$. If $t \in X^Y$ is such a tuple and $y \in Y$, then by $(t)_y$ we denote the $y$-th component of tuple $t$.

We now proceed to describe the formal syntax and semantics of our logical system starting with the definition of a transition system.

**Definition 1** A tuple $(W, \{\sim_a\}_{a \in \mathcal{A}}, V, M, \pi)$ is called an epistemic transition system, if

1. $W$ is a set of epistemic states,
2. $\sim_a$ is an indistinguishability equivalence relation on $W$ for each $a \in \mathcal{A}$,
3. $V$ is a nonempty set called domain of choices,
4. $M \subseteq W \times V^A \times W$ is an aggregation mechanism,
5. $\pi$ is a function that maps propositional variables into subsets of $W$.

For example, in the transition system $T_1$ depicted in Figure 1, the set of states $W$ is $\{w_0, w_1, w_2, w'_0, w'_1, w'_2\}$ and relation $\sim_a$ is a transitive reflexive closure of $\{(w_0, w'_0), (w'_1, w'_1), (w_2, w'_2)\}$.

Informally, an epistemic transition system is regular if there is at least one next state for each outcome of the vote.

**Definition 2** An epistemic transition system $(W, \{\sim_a\}_{a \in \mathcal{A}}, V, M, \pi)$ is regular if for each $w \in W$ and each $s \in V^A$, there is $w' \in W$ such that $(w, s, w') \in M$.

A coalition is a subset of $\mathcal{A}$. A strategy profile of coalition $C$ is any tuple in the set $V^C$.

**Definition 3** For any states $w_1, w_2 \in W$ and any coalition $C$, let $w_1 \sim_C w_2$ if $w_1 \sim_a w_2$ for each agent $a \in C$.

**Lemma 1** For each coalition $C$, relation $\sim_C$ is an equivalence relation on the set of epistemic states $W$.

**Definition 4** For all strategy profiles $s_1$ and $s_2$ of coalitions $C_1$ and $C_2$ respectively and any coalition $C \subseteq C_1 \cap C_2$, let $s_1 =_C s_2$ if $(s_1)_a = (s_2)_a$ for each $a \in C$.

**Lemma 2** For any coalition $C$, relation $=_C$ is an equivalence relation on the set of all strategy profiles of coalitions containing coalition $C$.

**Definition 5** A history is an arbitrary sequence $h = (w_0, s_1, w_1, s_2, w_2, \ldots, s_n, w_n)$ such that $n \geq 0$ and

1. $w_i \in W$ for each $i \leq n$,
2. $s_i \in V^A$ for each $i \leq n$,
3. $(w_i, s_{i+1}, w_{i+1}) \in M$ for each $i < n$.

In this paper we assume that votes of all agents are private. Thus, an individual agent only knows her own votes and the equivalence classes of the states that the system has been at. This is formally captured in the following definition of indistinguishability of histories by an agent.

**Definition 6** For any history $h = (w_0, s_1, w_1, \ldots, s_n, w_n)$, any history $h' = (w'_0, s'_1, w'_1, \ldots, s'_m, w'_m)$, and any agent $a \in \mathcal{A}$, let $h \approx_a h'$ if $n = m$ and

1. $w_i \sim_a w'_i$ for each $i \leq n$,
2. \((s_i)_a = (s'_i)_a\) for each \(i \leq n\).

**Definition 7** For any histories \(h_1, h_2\) and any coalition \(C\), let \(h_1 \approx_C h_2\) if \(h_1 \approx_a h_2\) for each agent \(a \in C\).

**Lemma 3** For any coalition \(C\), relation \(\approx_C\) is an equivalence relation on the set of histories.

The length \(|h|\) of a history \(h = (w_0, s_1, w_1, \ldots, s_n, w_n)\) is the value of \(n\). By Definition 7, the empty coalition cannot distinguish any two histories, even of different lengths.

**Lemma 4** \(|h_1| = |h_2|\) for each histories \(h_1, h_2\) such that \(h_1 \approx_C h_2\) for some nonempty coalition \(C\).

For any sequence \(x = x_1, \ldots, x_n\) and an element \(y\), by sequence \(x :: y\) we mean \(x_1, \ldots, x_n, y\). If sequence \(x\) is nonempty, then by \(hd(x)\) we mean element \(x_n\).

**Lemma 5** If \((h_1 :: s_1 :: w_1) \approx_C (h_2 :: s_2 :: w_2)\), then \(h_1 \approx_C h_2, s_1 \approx_C s_2,\) and \(w_1 \sim_C w_2\).

**Definition 8** Let \(\Phi\) be the language specified as follows:

- \(\varphi := p \mid \neg \varphi \mid \varphi \rightarrow \varphi \mid K_C \varphi \mid H_C \varphi\), where \(C \subseteq A\).

Boolean constants \(\perp, \top\) and \((=\) are defined as usual.

**Definition 9** For any history \(h\) of an epistemic transition system \((W, \{\alpha\}_\alpha \subseteq A, V, M, \pi)\) and any formula \(\varphi \in \Phi\), let satisfiability relation \(h \models \varphi\) be defined as follows:

1. If \(p \models \varphi\) if \(hd(h) \in \pi(p)\) and \(p\) is a propositional variable.
2. \(\models \neg \varphi\) if \(\models \varphi\) is false.
3. \(\models \varphi \rightarrow \psi\) if \(\models \varphi\) or \(\models \psi\).
4. \(\models K_C \varphi\) if \(\models K_C \psi\) for each history \(h\) s.t. \(h \approx_C h'\).
5. \(\models H_C \varphi\) if there is a strategy profile \(s \in V_C\) such that for any history \(h' :: s' :: w'\), if \(h \approx_C h'\) and \(s =_C s'\), then \(h' :: s' :: w' \models \varphi\).

**Axioms**

In additional to propositional tautologies in language \(\Phi\), our logical system consists of the following axioms:

1. Truth: \(K_C \varphi \rightarrow \varphi\).
2. Negative Introspection: \(\neg K_C \varphi \rightarrow K_C \neg K_C \varphi\).
3. Distributivity: \(K_C (\varphi \rightarrow \psi) \rightarrow (K_C \varphi \rightarrow K_C \psi)\).
4. Monotonicity: \(K_C \varphi \rightarrow K_D \varphi\), if \(C \subseteq D\).
5. Strategic Positive Introspection: \(H_C \varphi \rightarrow K_C H_C \varphi\).
6. Cooperation: \(H_C (\varphi \rightarrow \psi) \rightarrow (H_D \varphi \rightarrow H_C H_D \psi)\), where \(C \cap D = \emptyset\).
7. Empty Coalition: \(K_C \varphi \rightarrow H_C \varphi\).
8. Perfect Recall: \(H_D \varphi \rightarrow H_D K_C \varphi\), where \(D \subseteq C \neq \emptyset\).
9. Unachievability of Falsehood: \(\neg H_C \bot\).

We write \(\models \varphi\) if formula \(\varphi\) is provable from the axioms of our logical system using Necessitation, Strategic Necessitation, and Modus Ponens inference rules:

\[
\begin{array}{c|c|c}
\varphi & \varphi & \varphi \rightarrow \psi \\
K_C \varphi & H_C \varphi & \psi \\
\end{array}
\]

We write \(\models X \varphi\) if formula \(\varphi\) is provable from the theorems of our logical system and a set of additional axioms \(X\) using only Modus Ponens inference rule.

The next lemma follows from a well-known observation that the Positive Introspection axiom is provable from the other axioms of S5. The proof can be found in the full version of this paper (Naumov and Tao 2017b).

**Lemma 6** \(K_C \varphi \rightarrow K_C K_C \varphi\).

The proof of the following soundness theorem is also given in (Naumov and Tao 2017b).

**Theorem 1** If \(\models \varphi\), then \(\models \varphi\) for each history of each regular epistemic transition system.

**Derivation Examples**

This section contains examples of formal proofs in our logical system. The results obtained here are used in the proof of completeness. The proof of Lemma 7 is based on the one proposed to us by Natasha Alechina.

**Lemma 7 (Alechina)** \(\models \neg H_C \varphi \rightarrow K_C \neg H_C \varphi\).

**Proof.** By the Positive Strategic Introspection axiom, \(\models H_C \varphi \rightarrow K_C H_C \varphi\). Thus, \(\models \neg K_C H_C \varphi \rightarrow \neg H_C \varphi\) by the contrapositive. Hence, \(\models K_C (\neg K_C H_C \varphi \rightarrow \neg H_C \varphi)\) by the Necessitation inference rule. Then, by the Distributivity axiom and the Modus Ponens inference rule, \(\models K_C (\neg K_C H_C \varphi \rightarrow K_C \neg H_C \varphi)\). Thus, by the Negative Introspection axiom and the laws of propositional reasoning, \(\models \neg K_C H_C \varphi \rightarrow K_C \neg H_C \varphi\). Note that \(\neg H_C \varphi \rightarrow \neg K C H_C \varphi\) is the contrapositive of the Truth axiom. Therefore, by the laws of propositional reasoning, \(\models \neg H_C \varphi \rightarrow K_C \neg H_C \varphi\).

**Lemma 8** \(\models H_C \varphi \rightarrow H_D \varphi\), where \(C \subseteq D\).

**Proof.** Note that \(\varphi \rightarrow \varphi\) is a propositional tautology. Thus, \(\models \varphi \rightarrow \varphi\). Hence, \(\models H_D (\varphi \rightarrow \varphi)\) by the Strategic Necessitation inference rule. At the same time, by the Cooperation axiom, \(\models H_D (\varphi \rightarrow \varphi) \rightarrow (H_C \varphi \rightarrow H_D \varphi)\) due to the assumption \(C \subseteq D\). Therefore, \(\models H_C \varphi \rightarrow H_D \varphi\) by the Modus Ponens inference rule.

**Lemma 9** If \(\varphi_1, \ldots, \varphi_n \models \psi\), then

1. \(K_C \varphi_1, \ldots, K_C \varphi_n \models K_C \psi\),
2. \(H_C \varphi_1, \ldots, H_C \varphi_n \not\models H_D \psi, \) where sets \(C_1, \ldots, C_n\) are pairwise disjoint.

**Proof.** To prove the second statement, apply deduction lemma for propositional logic \(n\) times. Then, we have \(\models \varphi_1 \rightarrow (\cdots (\varphi_n \rightarrow \psi))\). Thus, by the Strategic Necessitation inference rule, \(\models H_D (\varphi_1 \rightarrow (\cdots ((\varphi_n \rightarrow \psi)))\). Hence, \(\models H_C \varphi_1 \rightarrow H_C (\varphi_2 \rightarrow (\cdots ((\varphi_n \rightarrow \psi)))\) by the Cooperation axiom and the Modus Ponens inference rule. Then, \(\models \varphi_1 \rightarrow H_C (\varphi_2 \rightarrow (\cdots ((\varphi_n \rightarrow \psi)))\) by the Modus Ponens inference rule. Thus, again by the Cooperation axiom and the Modus Ponens inference rule we have \(H_C \varphi_1 \rightarrow H_C \varphi_2 \rightarrow H_C (\varphi_3 \rightarrow (\cdots ((\varphi_n \rightarrow \psi)))\). Therefore, by repeating the last two steps \(n - 2\) times, \(\models \varphi_1 \rightarrow H_C \varphi_2 \rightarrow H_C \varphi_3 \rightarrow (\cdots (\varphi_n \rightarrow \psi))\). The proof of the first statement is similar, but it uses the Distributivity axiom instead of the Cooperation axiom.
Completeness

In the rest of this paper we focus on the completeness theorem for our logical system with respect to the class of regular epistemic transition systems. We start the proof of completeness by fixing a maximal consistent set of formulae $X_0$ and defining a canonical epistemic transition system $ETS(X_0) = (W, \{\sim_a\}_{a \in A}, \Phi, M, \pi)$ using the “unravelling” technique (Sahlqvist 1975). Note that the domain of choices in the canonical model is the set of all formulae $\Phi$.

Canonical Epistemic Transition System

**Definition 10** The set of epistemic states $W$ consists of all sequences $X_0, C_1, X_1, \ldots, C_n, X_n$, such that $n \geq 0$ and
1. $X_i$ is a maximal consistent subset of $\Phi$ for each $i \geq 1$,
2. $C_i \subseteq A$ for each $i \geq 1$,
3. $\{\varphi | K_C \varphi \in X_{i-1}\} \subseteq X_i$, for each $i \geq 1$.

**Definition 11** Suppose that $w = X_0, C_1, X_1, \ldots, C_n, X_n$ and $w' = X_0, C_1', X_1', \ldots, C_n', X_n'$ are epistemic states. For any agent $a \in A$, let $w \sim_a w'$ if there is a non-negative integer $k \leq \min\{n, m\}$ such that
1. $X_i = X_i'$ for each $i$ such that $0 < i \leq k$,
2. $C_i = C_i'$ for each $i$ such that $0 < i \leq k$,
3. $a \in C_i$ for each $i$ such that $k < i \leq n$,
4. $a \in C_i'$ for each $i$ such that $k < i \leq m$.

The next lemma states the basic property of the indistinguishability relation $\sim_a$. The proof of this lemma is standard. See (Naumov and Tao 2017b) for details.

**Lemma 10** For any epistemic states $w, w' \in W$ such that $w \sim C w'$, if $K_C \varphi \in hd(w)$, then $\varphi \in hd(w')$.

Next, we specify the aggregation mechanism of the canonical epistemic transition system. Informally, if an coalition has a know-how strategy to achieve $\varphi$ and all members of the coalition vote for $\varphi$, then $\varphi$ must be true after the transition.

**Definition 12** For any states $w, w' \in W$ and any complete strategy profile $s \in \Phi^A$, let $(w, s, w') \in M$ if
$$\{\varphi | (H_D \varphi \in hd(w)) \wedge \forall a \in D((s)_a = \varphi)\} \subseteq hd(w').$$

**Definition 13** $\pi(p) = \{w | w \in W | p \in hd(w)\}$.

This concludes the definition of the canonical epistemic transition system $ETS(X_0) = (W, \{\sim_a\}_{a \in A}, \Phi, M, \pi)$. We prove that this system is regular in Lemma 18.

K-child Lemmas

The following technical results (Lemmas 11–15) about the knowledge modality $K$ are used in the proof of completeness.

**Lemma 11** For any epistemic state $w \in W$ if $\neg K_C \varphi$ is in $hd(w)$, then there is an epistemic state $w' \in W$ such that $w \sim C w'$ and $\neg \varphi \in hd(w')$.

**Proof.** Consider the set $X = \{\neg \varphi\} \cup \{\psi | K_C \psi \in hd(w)\}$. First, we show that set $X$ is consistent. Assume the opposite. Then, there exist formulae $K_C \psi_1, \ldots, K_C \psi_n \in hd(w)$ such that $\psi_1, \ldots, \psi_n \vdash \neg \varphi$. Thus, $K_C \psi_1, \ldots, K_C \psi_n \vdash K_C \varphi$ by Lemma 9. Therefore, $hd(w) \vdash K_C \varphi$ by the choice of formulae $K_C \psi_1, \ldots, K_C \psi_n$, which contradicts the consistency of the set $hd(w)$ due to the assumption $\neg K_C \varphi \in hd(w)$.

Let $\hat{X}$ be a maximal consistent extension of set $X$ and let $w'$ be sequence $w :: C :: X$. Note that $w' \in W$ by Definition 10 and the choice of set $X$. Furthermore, $w \sim C w'$ by Definition 11. To finish the proof, note that $\neg \varphi \in X \subseteq \hat{X} = hd(w')$ by the choice of set $X$.

History $h$ is a sequence whose last element $hd(h)$ is an epistemic state. Epistemic state $hd(h)$, by Definition 10, is also a sequence. Expression $hd(h)$ denotes the last element of the sequence $hd(h)$.

**Lemma 12** For any history $h$, if $K_C \varphi \in hd(h)$, then $\varphi \in hd(h')$ for each history $h'$ such that $h \approx C h'$.

**Proof.** Assumption $h \approx C h'$ by Definition 6 implies that $hd(h) \sim C hd(h')$. Therefore, $\varphi \in hd(h'(h))$ by Lemma 10 and the assumption $K_C \varphi \in hd(h)$.

Consider the set of formulae $X = \{\varphi | K_C \varphi \in hd(w_1)\} \cup \{\neg H_D \psi | \neg \psi \in hd(w_3) \wedge \forall a \in D((s')_a = \psi)\}$. First, we show that set $X$ is consistent. Indeed, set $\neg H_D \psi | \neg \psi \in hd(w_3) \wedge \forall a \in D((s')_a = \psi)$ is equal to the union of the following two sets
$$\{\neg H_D \psi | \neg \psi \in hd(w_3) \wedge D \subseteq C \wedge \forall a \in D((s')_a = \psi)\},$$
$$\{\neg H_D \psi | \neg \psi \in hd(w_3) \wedge D \subseteq C \wedge \forall a \in D((s')_a = \psi)\}.$$
Suppose the opposite. In other words, assume that there are formulae $K_C \varphi_1, \ldots, K_C \varphi_n \in hd(w_1)$, \hfill (2)  
and sets $D_1, \ldots, D_m \subseteq C$, \hfill (3)  
such that \hfill (4) 
and \hfill (5) 

By Lemma 9 and the Truth axiom,

\[ K_C \varphi_1, \ldots, K_C \varphi_n, K_C \neg H_{D_1} \psi_1, \ldots, K_C \neg H_{D_m} \psi_m \vdash \bot. \]

Hence, statement (2) and the consistency of the set $hd(w_1)$ imply that there exists $k \leq m$ such that $K_C \neg H_{D_k} \psi_k \not\in hd(w_1)$. Thus, $K_C \neg H_{D_k} \psi_k \in hd(w_1)$ due to the maximality of the set $hd(w_1)$. Then, $K_{D_k} \neg H_{D_k} \psi_k \not\in hd(w_1)$ by statement (4) and the contrapositive of the Monotonicity axiom. Then, $hd(w_1) \supseteq D_k \psi_k$ by the contrapositive of Lemma 7. Thus, $hd(w_1) \supseteq H_{D_1} K_C \psi_k$ by the Perfect Recall axiom, statement (4), and the assumption of the lemma that $C \not\in \emptyset$. Hence, $H_{D_1} K_C \psi_k \in hd(w_1)$ due to the maximality of the set $hd(w_1)$. Note that statements (1) and (4) imply $s = _{D_k} s'$. Then, $(s)_a = \psi_k$ for each agent $a \in D_k$ by statement (5). Thus, $K_C \psi_k \in hd(w_2)$ by assumption $(w_1, s, w_2) \in M$, statement (4), and Definition 12. Hence, $\psi_k \in hd(w_3)$, by the assumption $w_2 \sim C w_3$ of the lemma and Lemma 10. This contradicts statement (3) and the consistency of the set $hd(w_3)$. Therefore, set $X$ is consistent.

Let $X$ be any maximal consistent extension of set $X$. Define $w_4$ to be the sequence $w_1 :: C :: X$. Note that $w_4 \in W$ by Definition 10 and the choice of set $X$. At the same time, $w_1 \sim_C w_4$ by Definition 11 and Definition 3.

Finally, let us show that $(w_4, s', w_3) \in M$ using Definition 12. Consider any $D \subseteq A$ and any $H_D \psi \in hd(w_4) = X$ such that $s'(a) = \psi$ for each $a \in D$. We need to show that $\psi \in hd(w_3)$. Suppose the opposite. Then, $\neg \psi \in hd(w_3)$ by the maximality of set $hd(w_3)$. Thus, $\neg H_{D'} \psi \not\in X$ by the choice of set $X$. Hence, $H_{D'} \psi \not\in X \subseteq X$. Therefore, $H_{D'} \psi \not\in X$ by the consistency of set $X$, which contradicts the choice of formula $H_D \psi$. ☒

\textbf{Lemma 14} For any history $h$, if $K_C \varphi \not\in hd(h(h))$, then there is a history $h'$ s.t. $h \approx_C h'$ and $\neg \varphi \in hd(h(h'))$.

\textbf{Proof.} By Lemma 11, there is a state $w \in W$ such that $hd(h) \sim_C w$ and $\neg \varphi \in hd(w)$. Let $h'$ be a one-element sequence $w$. Note that $h \approx_C h'$ by Definition 7. Finally, $\neg \varphi \in hd(w) = hd(h(h'))$. ☒

\textbf{Lemma 15} For any nonempty coalition $C$ and any history $h$, if $K_C \varphi \not\in hd(h(h))$, then there is a history $h'$ such that $h \approx_C h'$ and $\neg \varphi \in hd(h(h'))$.

\textbf{Proof.} Let $h = (w_0, s_1, w_1, \ldots, s_n, w_n)$. We prove the lemma by induction on integer $n$.

\textbf{Base Case.} Let $n = 0$. By Lemma 11, there is $w'_0 \in W$ such that $w_0 \sim_C w'_0$ and $\neg \varphi \in hd(w'_0)$. Let $h'$ be a one-element sequence $w'_0$. Note that $w_0 \sim_C w'_0$ implies that $h \approx_C h'$ by Definition 6. Also, $\neg \varphi \in hd(w'_0) = hd(h(h'))$.

\textbf{Induction Step.} Let $h = (w_1, s_2, \ldots, s_n, w_n)$, see Figure 4. By Definition 5, sequence $h$ is a history. By the induction hypothesis there is a history $h'$ such that $h \approx_C h'$ and $\neg \varphi \in hd(h(h'))$. Let $h'_1 = (w'_1, s'_2, \ldots, s'_n, w'_n)$. By Lemma 13, there is a state $w'_0$ and a complete strategy profile $s'_0$ such that $w_0 \sim_C w'_0$ and $(s'_0, s'_1) \in M$, and $s_1 = s'_1$.

Define $h' = (w'_0, s'_1, w'_1, s'_2, \ldots, s'_n, w'_n)$. By Definition 5, statement $(w'_0, s'_1, w'_1) \in M$ implies that $h'$ is a history. Note that $h \approx_C h'$ by Definition 6 because $h'_1 \approx_C h_1$, $w_0 \sim_C w'_0$, and $s_1 = s'_1$. Finally, $\neg \varphi \in hd(h'(h'))$ because $hd(h'(h')) = hd(h(h'))$ and $\neg \varphi \in hd(h(h'))$. ☒

\textbf{H-child Lemmas} Lemmas 16 and 17 are about the knowhow modality $H$. They are used later in the proof.

\textbf{Lemma 16} For any history $h$, if $H_C \varphi \in hd(h(h))$, then there is a strategy profile $s$ of coalition $C$ s.t. for any history $h' :: s' :: w'$ if $h \approx_C h'$ and $s = C s'$, then $\varphi \in hd(w')$.

\textbf{Proof.} Consider a strategy profile $s$ of coalition $C$ such that $(s)_a = \varphi$ for each agent $a \in C$. Suppose that $H_C \varphi \in hd(h(h))$ and sequence $h' :: s' :: w'$ is a history such that $h \approx_C h'$ and $s = C s'$. It suffices to show that $\varphi \in hd(w')$.

By the Strategic Positive Introspection axiom, assumption $H_C \varphi \in hd(h(h))$ implies $hd(h(h)) \supseteq K_C H_C \varphi$. Thus, $K_C H_C \varphi \in hd(h(h))$ by the maximality of set $hd(h(h))$.

Assumption $h \approx_C h'$ implies $hd(h) \sim_C hd(h')$ by Definition 6. Thus, $H_C \varphi \in hd(h(h'))$ by Lemma 10 and because $K_C H_C \varphi \in hd(h(h))$.

By Definition 5, assumption that sequence $h' :: s' :: w'$ is a history implies that $(hd(h'), s', w') \in M$. Thus, $\varphi \in hd(w')$ by Definition 12 because $H_C \varphi \in hd(h(h'))$ and $(s')_a = (s)_a = \varphi$ for each agent $a \in C$. ☒

\textbf{Lemma 17} For any history $h$ and any strategy profile $s$ of a coalition $C$, if $\neg H_C \varphi \in hd(h(h))$, then there is a history $h :: s' :: w'$ such that $s = C s'$ and $\neg \varphi \in hd(w')$.

\textbf{Proof.} Let $w = hd(h)$. Consider a complete strategy profile $s'$ and a set of formulæ $X$ such that

\[ (s')_a = \begin{cases} (s)_a, & \text{if } a \in C, \\ T, & \text{otherwise}, \end{cases} \]  

\hfill (6)
Lemma 18 \textit{The systemETS}(X_0) is regular}.

\textbf{Proof.} Let \(w \in W\) and \(s \in V^A\). By Definition 2, it suffices to show that there is an epistemic state \(w'\) such that \((w, s, w') \in M\). Indeed, let history \(h\) be a single-element sequence \(w\). Note that \(h_{A} \subseteq \text{hd}(\text{hd}(h))\) by the Unachievability of Falsehood axiom and due to the maximality of set \(\text{hd}(w)\). Thus, by Lemma 17, there is a history \(h' \models A s'\) such that \(s = A s'\). Hence, \((\text{hd}(h), s', w') \in M\) by Definition 5. At the same time, \(s = A s'\) implies that \(s = s'\) by Definition 4. Thus, \((\text{hd}(h), s, w') \in M\). Therefore, \((w, s, w') \in M\) because \(\text{hd}(h) = w\).

\textbf{Completeness: Final Steps} \hfill \\

\textbf{Lemma 19} \(h \models \varphi\) iff \(\varphi \in \text{hd}(\text{hd}(h))\) for each history \(h\) and each formula \(\varphi \in \Phi\).

\textbf{Proof.} We prove the statement by induction on the structural complexity of formula \(\varphi\). If \(\varphi\) is an atomic proposition \(p\), then \(h \models p\) iff \(\text{hd}(h) \in \pi(p)\), by Definition 9. Hence, \(h \models p\) iff \(p \in \text{hd}(\text{hd}(h))\) by Definition 13.

The cases when formula \(\varphi\) is a negation or an implication follow from Definition 9 and the maximality and the consistency of the set \(\text{hd}(\text{hd}(h))\) in the standard way.

Next, suppose that formula \(\varphi\) has the form \(K_C \psi\).

\(\models \psi \in \text{hd}(\text{hd}(h))\). Then, either by Lemma 14 (when set \(C\) is empty) or by Lemma 15 (when set \(C\) is nonempty), there is a history \(h'\) such that \(h \models C h'\) and \(\psi \in \text{hd}(\text{hd}(h'))\). Hence, \(h' \not\models \psi\) by the induction hypothesis. Therefore, \(h \not\models K_C \psi\) by Definition 9.

\(\models \psi \in \text{hd}(\text{hd}(h))\). Then, there is a history \(h'\) such that \(h \models C h'\) and \(h' \not\models \psi\). Thus, \(\psi \not\in \text{hd}(\text{hd}(h'))\) by the induction hypothesis. Then, \(K_C \varphi \not\in \text{hd}(\text{hd}(h))\) by Lemma 12.

Finally, let formula \(\varphi\) be of the form \(H_C \psi\).

\(\models \psi \in \text{hd}(\text{hd}(h))\). Then, by Definition 9, there is a strategy profile \(s \in C^G\) such that for any history \(h': s' \models w'\) if \(h \models C h'\) and \(s = C s'\), then \(h' \models s' \models w' \models \psi\). Thus, by Lemma 3,

\begin{equation}
\text{for any history } h \models s' \models w', \text{ if } s = C s', \text{ then } h \models s' \models w' \models \psi.
\end{equation}

Suppose that \(H_C \psi \not\in \text{hd}(\text{hd}(h))\). Then, \(H_C \psi \not\in \text{hd}(\text{hd}(h))\) due to the maximality of the set \(\text{hd}(\text{hd}(h))\). Hence, by Lemma 17, there is a history \(h \models s' \models w'\) such that \(s = C s'\) and \(\psi \not\in \text{hd}(w')\). Thus, \(\psi \not\in \text{hd}(w')\) due to the consistency of set \(\text{hd}(w')\). Hence, by the induction hypothesis, \(h \models s' \models w' \not\models \psi\), which contradicts statement (11).

\(\models \psi \in \text{hd}(\text{hd}(h))\). By Lemma 16, there is a strategy profile \(s \in C^G\) such that for any history \(h' \models s' \models w'\) if \(h \models C h'\) and \(s = C s'\), then \(\psi \in \text{hd}(w')\). Hence, by the induction hypothesis, for any history \(h' \models s' \models w'\) if \(h \models C h'\) and \(s = C s'\), then \(h' \models s' \models w' \models \psi\). Therefore, \(h \models H_C \psi\) by Definition 9.

\textbf{Theorem 2} \(h \models \varphi\) for each history \(h\) of each regular epistemic transition system, \(\models \varphi\).

\textbf{Proof.} Suppose that \(\models \varphi\). Consider any maximal consistent set \(X_0\) such that \(\varphi \notin X_0\). Let \(h_0\) be a single-element sequence consisting of just set \(X_0\). By Definition 5, sequence \(h_0\) is a history of the canonical epistemic transition system \(ETS(X_0)\). Then, \(h_0 \models \varphi\) by Lemma 19. Therefore, \(h_0 \not\models \varphi\) by Definition 9.

Theorem 2 can be generalized to the strong completeness theorem in the standard way. We also believe that the number of states and the domain of choices in the canonical model can be made finite using filtration on subformulas.

\textbf{Conclusion}

We have extended the recent study of the interplay between knowledge and strategic coalition power (Ágotnes
References


