Prevailing in the Dark: Information Walls in Strategic Games

Pavel Naumov,[1] **Wenxuan Zhang** [2]

[1] University of Southampton, Southampton, United Kingdom
[2] Scripps College; Claremont, California, United States
p.naumov@soton.ac.uk, wzhang4343@scrippscollege.edu

Abstract

The paper studies strategic abilities that rise from restrictions on the information sharing in multiagent systems. The main technical result is a sound and complete logical system that describes the interplay between the knowledge and the strategic ability modalities.

Introduction

Controlled by al-Hazmi and al-Mihdhar, American Airline flight 77 crashed into the Pentagon at 9:37:46 am on September 11th, 2001 (Kean 2004). The months that preceded this event could be viewed as a strategic game between Al-Qaeda headquarters, U.S. Federal Bureau of Investigation (FBI), and U.S. Central Intelligence Agency (CIA). As later investigation revealed, CIA had evidence that al-Hazmi and al-Mihdhar were Al-Qaeda agents living in California and potentially plotting an attack on the United States. However, American laws prevented CIA from arresting them on American land. FBI, whose informant lived under the same roof with al-Hazmi and al-Mihdhar, suspected that they were a potential threat, but did not have enough evidence to open a preliminary inquiry to investigate them. If CIA would have shared their knowledge with FBI, the latter could have arrested them and, thus, prevented the attack on Pentagon. CIA did not share this information with FBI[1] because it believed that it was illegal to share information between intelligence and criminal investigations of a common target (Grewe 2004). As a result, al-Hazmi and al-Mihdhar were left free to live in California and to high-jack the plane once ordered to do so by Al-Qaeda headquarters.

Figure 1 captures the above situation as a strategic game between three agents: Al-Qaeda, FBI, and CIA. This game has two "initial" states, w and w', and two "final" states, u and v. In state w, al-Hazmi and al-Mihdhar are Al-Qaeda agents preparing for the attack on the United States. In state w', they are peaceful foreigners taking flight lessons in San Diego. The actual initial state of the game was w. CIA was able to distinguish this state from state w', but FBI was not. We show the indistinguishability relation by dashed line in

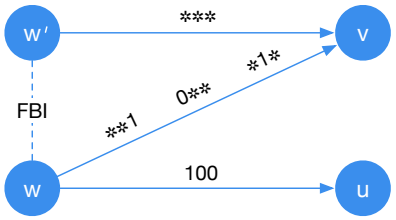

Figure 1: Strategic game between Al-Qaeda, FBI, and CIA

Figure 1. Final state u is the one in which flight 77 crashes into the Pentagon. Final state v is when it does not.

We assume that each of the three agents has two actions available in state w and w': action 0 and action 1. For CIA, action 1 means "arrest" and action 0 means "do not arrest" al-Hazmi and al-Mihdhar. Because it is illegal for CIA to conduct operations on American land, action 1 is illegal for CIA in state w as well as in state w'. For FBI, action 1 means "open an investigation" and action 0 means "do not open the investigation." FBI cannot conduct searches or electronic surveillance without probable cause unless the target is an agent of foreign power. Thus, it would be proper for FBI to conduct an investigation in state w but not in state w'. For Al-Qaeda, action 1 means "order the attack", action 0 means "do not order the attack."

In this game, an action profile is a tuple afc, where a, f, and c are actions of Al-Queda, FBI, and CIA, respectively. The directed edges in Figure 1 show transitions of the strategic game from an initial to a final state under different action profiles. For example, edge from state w' to state v is labeled with $***$. This means that the game transitions from state w' to state v under *any* action profile afc. At the same time, directed edge from state w to state u is labeled with 100. It means that this transition can only happen if Al-Queda decides to attack and FBI and CIA decide not to act.

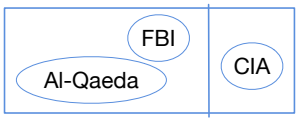

Figure 2: Information wall defining partition σ.

[1] CIA placed al-Hazmi and al-Mihdha on FBI's watch list in August 2001, but this did not give FBI enough time to investigate.

The key to understanding this strategic game is an *information wall* that existed between FBI and CIA. In this paper, we model such walls as partitions of the set of all agents. One can consider a single-wall partition depicted in Figure 2 with Al-Qaeda and FBI on one side of the wall and CIA on the other. One can also consider a partition where Al-Qaeda is moved to the CIA side. Finally, it is possible to introduce a two-wall partition that prevents communication between all three agents. All such cases could be modeled using the formalism that we propose in this paper, but for the sake of this example, we assume only the information wall shown in Figure 2. We refer to the partition defined by this wall as partition σ. If this wall did not exist, CIA and FBI might share the information or might not. The existence of the wall gave Al-Qaeda a *strategy* that guaranteed Al-Qaeda's ability to achieve its goal. We denote the existence of such a strategy by

$$\mathsf{H}^{\sigma}_{\text{Al-Qaeda}}(\text{“flight 77 crashed into the Pentagon”}).$$

In general, we write $\mathsf{H}^{\sigma}_{a}\varphi$ if agent a has a strategy to achieve φ as long as there are information walls in the game defined by the partition σ. As common in the games with imperfect information, by a strategy we mean a know-how (or uniform) strategy. A strategy of an agent a is a know-how strategy if agent a can use the same strategy to achieve the desired result from each state indistinguishable by the agent from the current state. In other words, a strategy is a know-how strategy if the strategy exists, the agent knows that it exists, and the agent also knows what the strategy is.

Our next example models a very simplified version of media censorship. In this example, each of the three agents Luo Ji, Zhi Shi, and Tui Li can be either content (C) or discontent (D) with the current political situation. As a result, the game has eight initial states labeled CCC, ..., DDD in Figure 3. For example, in state CDC Luo Ji is content (C), Zhi Shi is discontent (D), and Tui Li is content (C). Each of the agents knows if she herself is content or discontent, but does not know the other two agents' positions. Thus, for example, Luo Ji can distinguish states DDC from state CCD, but not state DDC from state DCC. Dashed lines between initial states in Figure 3 connect states indistinguishable by Luo Ji.

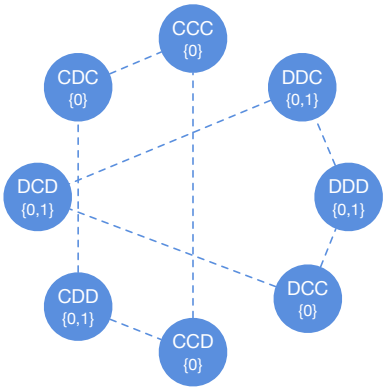

Figure 3: Initial states for censorship example.

Each of the three agents has two available actions in each initial state: to remain silent (action 0) or to question the government (action 1). If at least one of the agents questions the government and the majority of people is discontent, then a revolution happens and none of the agents is prosecuted. If an agent questions the government but majority is content, then no revolution happen and the questioning agent is executed. Finally, if none of the agents questions the government, then the status quo remains. Each agent first and foremost does not want to be executed. We call an action that does not result in the agent's execution a *safe* action of the agent. Safe actions were called "legal" in our previous example. In general, by a safe action we mean any action that satisfied certain external or self-imposed constraints on the agents. In state DDC both actions, 0 and 1, are safe for each agent. In state DCC only action 0 is safe for each agent. In Figure 3, each state is labeled by the set of safe actions. In our example, the set of safe actions for each agent is the same in a given state. In general, we allow the set of safe actions to be agent-specific. Let S^{w}_{a} denote the set of all safe actions for agent a in state w. Thus, $S^{DDC}_{\text{Luo Ji}} = \{0, 1\}$ and $S^{DCC}_{\text{Luo Ji}} = \{0\}$.

Note that for all three agents action 1 is safe in state DDC, but not in state DCC. Suppose that Luo Ji and Zhi Shi are discontent and Tui Li is not. Thus, the actual state is DDC. Since Luo Ji cannot distinguish states DDC and DCC, he does not *know* that the action 1 is safe. If an agent knows that an action is safe, then we say that the action is *knowingly safe* for the agent. By KS^{w}_{a} we denote the set of all knowingly safe actions for agent a in state w. Then, $KS^{DDC}_{\text{Luo Ji}} = KS^{DCC}_{\text{Luo Ji}} = \{0\}$.

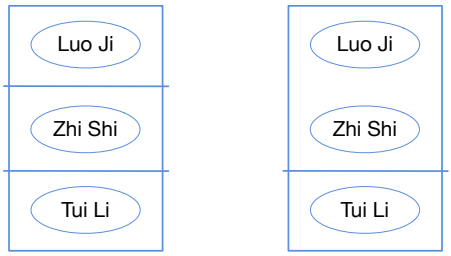

Figure 4: Partitions τ_1 (left) and τ_2 (right).

Suppose that there is an information wall that prevents any communication between agents Luo Ji, Zhi Shi, and Tui Li. We identify this wall with partition τ_1 on the set of all agents, see Figure 4 (left). Furthermore, recall that the current state is DDC. In other words, the majority is discontent. Then, it is safe for each agent to question the government, but none of them will ever learn this because information wall τ_1 is present. Thus, none of the agents will question the government, and it is guaranteed that the revolution will not happen. We write this as

$$\mathsf{N}^{\tau_1}(\text{“no revolution happened”}).$$

In general, we write $\mathsf{N}^{\tau}\varphi$ if statement φ is guaranteed ("nesssary") to be true when information wall τ is present.

Let us consider now the case, see Figure 4 (right), when the information wall only separates Tui Li from the other

two agents. In this case, Luo Ji *might* communicate with Zhi Shi and learn that Zhi Shi is also discontent. Then, Luo Ji will conclude that action 1 is safe and he *might* start the revolution by questioning the government. In other words, information wall defined by partition τ_2 does not guarantee that the revolution will not happen:

$$\neg \mathsf{N}^{\tau_2}(\text{“no revolution happened”}).$$

Another example of information walls is a common business practice[2] to either prohibit or to discourage discussions of salaries at the workplace (Hayes 2017). The goal of this practice is to discourage employees from negotiating for higher salary.

In this paper we study the interplay between modalities H^{σ}_{a}, N^{σ}, and the individual knowledge modality K_a.

Related Literature

This work combines two previously independent lines of research: study of information flow for a given communication network topology and the study of agents' and their coalitions' strategic abilities.

Logics of Information Flow

In the existing literature, restrictions on the information exchange between agents are usually imposed by specifying a *graph* of the communication channels between the agents.

Pacuit and Parikh proposed a logical system combining knowledge modality with a communication modality over edges of a given graph (2004; 2007). They do not prove completeness of their system. More and Naumov (2011b; 2011a) gave a complete logical system for reasoning about Sutherland's nondeducibility relation (1986) between communication channels in a given graph and hypergraph. They treated nondeducibility relation as an atomic proposition, not as a modality. Donders, More, and Naumov did the same for directed acyclic graphs (2011).

Kane and Naumov proposed a sound and complete epistemic logic for reasoning about information flow on a linear graph (2013). Naumov and Tao generalized it to an arbitrary graph (2017b).

The set of connected components of any undirected graph defines a partition of the set of all agents into groups. The agents in different groups are not able to communicate. In this paper we represent restrictions on communication between agents through a partition rather than a graph. We do this because the exact structure of communication channels in each connected component is not significant for our work and also because partitions allow us to state axioms of our logical system in a more succinct and elegant form.

Logics of Strategic Abilities

Logics of coalition power were developed by Pauly, who also proved the completeness of the basic logic of coalition power (2001; 2002). Pauly's approach has been widely studied in the literature (Goranko 2001; van der Hoek and Wooldridge 2005; Borgo 2007; Sauro et al. 2006; Ågotnes et al. 2010; Ågotnes, van der Hoek, and Wooldridge 2009; Belardinelli 2014). An alternative logical system was proposed by More and Naumov (2012). Alur, Henzinger, and Kupferman introduced Alternating-Time Temporal Logic (ATL) that combines temporal and coalition modalities (2002). Van der Hoek and Wooldridge proposed to combine ATL with epistemic modality to form Alternating-Time Temporal Epistemic Logic (2003). Goranko and van Drimmelen gave a complete axiomatization of ATL (2006). Decidability and model checking problems for ATL-like systems have also been widely studied (Aminof et al. 2016; Berthon et al. 2017; Berthon, Maubert, and Murano 2017). Wang proposed a complete axiomatization of "knowing how" as a binary modality (2015; 2018). An alternative approach to expressing the power to achieve a goal in a temporal setting is the STIT logic (Belnap and Perloff 1990; Horty and Belnap 1995; Horty 2001; Horty and Pacuit 2017; Olkhovikov and Wansing 2019). Broersen, Herzig, and Troquard have shown that the coalition logic can be embedded into a variation of STIT logic (2007). Another approach to reasoning about strategies is Strategy Logic (Chatterjee, Henzinger, and Piterman 2010; Mogavero et al. 2014; Berthon et al. 2017; Aminof et al. 2018) that introduces explicit quantifiers over strategies.

Several modal logical systems that capture the interplay between knowledge and know-how strategies have been proposed. Ågotnes and Alechina introduced a complete axiomatization of an interplay between single-agent knowledge and coalition know-how modalities to achieve a goal in one step (2019). Naumov and Tao proposed a modal logic that combines the distributed knowledge modality with the coalition know-how modality to maintain a goal (2017a). Fervari, Herzig, Li, and Wang developed a sound and 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 (2017). Naumov and Tao introduced a trimodal logical system that describes an interplay between the (not know-how) coalition strategic modality, the coalition know-how modality, and the distributed knowledge modality (2018c). They also proposed logical systems describing the properties of know-how strategies for perfect recall setting (2018b), for second-order know-how (2018a), and for know-how modality in a metric space (2019).

Our Contribution

The above papers on information flow deal with knowledge, but not strategic abilities. The cited papers on strategic abilities do not take into account constraints on communication between the agents. In this paper we propose a logical system that considers strategic abilities and constraints on agents' behavior that rise from restrictions on the information shared between agents.

Our main technical result is the completeness of the proposed logical system. The proof of the completeness uses two key ideas: σ-harmony and distributed key generation. While σ-harmony is a variation of the technique from (Naumov and Tao 2018c,a), distributed key generation is a novel

[2]This practice is illegal in the US under National Labor Relations Act of 1935.

technique that we propose. We are not aware of this technique being used in completeness proofs before although it is well-known in cryptography (Pedersen 1991).

This paper is organized as follows. The next section defines the syntax and the semantics of our logical system. Then, we list and discuss its axioms and inference rules. The soundness of these axioms is shown in the section Soundness. Afterwards, we explain the key ideas behind the proof of the completeness. Details of the completeness proof are given in the appendix.

Syntax and Semantics

In this paper, we assume a fixed set of propositional variables and a set $\mathcal{A}$ of agents. A partition of set $\mathcal{A}$ is any family of pairwise disjoint nonempty sets whose union is equal to $\mathcal{A}$. For any agent $a \in \mathcal{A}$ and any partition σ of set $\mathcal{A}$, let $[a]_\sigma$ be the unique set in partition σ that contains a. In Figure 4 (right), for instance, $[\text{Luo Ji}]_{\tau_2} = \{\text{Luo Ji},\ \text{Zhi Shi}\}$.

By σ/a we mean a modification of partition σ in which set $[a]_\sigma$ is replaced by two sets: $\{a\}$ and $[a]_\sigma \setminus \{a\}$. If $[a]_\sigma = \{a\}$, then σ/a, by definition, is σ. For example,

$$\begin{aligned} \tau_2/\text{Luo Ji} &= \{\{\text{Luo Ji}\}, \{\text{Zhi Shi}\}, \{\text{Tui Li}\}\} = \tau_1, \\ \tau_1/\text{Luo Ji} &= \tau_1. \end{aligned}$$

Definition 1. *For any two partitions σ and τ of set $\mathcal{A}$, let $\sigma \preceq \tau$ if $[a]_\sigma \subseteq [a]_\tau$ for each agent $a \in \mathcal{A}$. If $\sigma \preceq \tau$, then partition σ is "finer" than partition τ.*

The language Φ of our logical system is defined by the following grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{K}_a\varphi \mid \mathsf{N}^\sigma\varphi \mid \mathsf{H}^\sigma_a\varphi,$$

where p is a propositional variable, $a \in \mathcal{A}$ is an agent, and σ is a partition of the set of agents. We read $\mathsf{K}_a\varphi$ as "agent a knows φ", $\mathsf{H}^\sigma_a\varphi$ as "agent a knows a safe strategy to achieve φ in the presence of the information walls defined by partition σ", and $\mathsf{N}^\sigma\varphi$ as "φ is unavoidable (necessary) in the presence of the information walls defined by partition σ". We assume that conjunction $\wedge$ and Boolean constant true $\top$ are defined through $\to$ and $\neg$ in the standard way. For any finite set of formulae Y, by $\wedge Y$ we mean the conjunction of all formulae in Y. Formula $\wedge\varnothing$, by definition, is $\top$.

We define semantics of our logical system in terms of "games". In the definition below, $\Delta^{\mathcal{A}}$ is the set of all functions from set of agents $\mathcal{A}$ to domain of actions Δ. Such functions will be called "action profiles".

Definition 2. *A game is a tuple*

$$(W, \{\sim_a\}_{a\in\mathcal{A}}, \Delta, \{S^w_a\}^{w\in W}_{a\in\mathcal{A}}, \ell, M, \pi),\ \textit{where}$$

1. *W is a (possibly empty) set of "states",*
2. *$\sim_a$ is an "indistinguishability" equivalence relation on the set of states W, for each agent $a \in \mathcal{A}$,*
3. *Δ is a set called "domain of actions",*
4. *$S^w_a \subseteq \Delta$ is a set of "safe" actions for agent $a \in \mathcal{A}$ in state $w \in W$,*
5. *$\ell \in S^w_a$ is a "default" action, which is safe for each agent $a \in \mathcal{A}$ in each state $w \in W$,*
6. *$M \subseteq W \times \Delta^{\mathcal{A}} \times W$ is a "mechanism" relation,*
7. *$\pi(p) \subseteq W$ for each propositional variable p.*

In our censorship example, set W includes "initial" states CCC, CCD, ..., DDD as well as different "outcome" states corresponding to revolution and no revolution with and without execution of agents. The outcome states are not depicted in Figure 3. Indistinguishability relation on initial states for agent Luo Ji is captured by dashed lines in Figure 3. The set of actions Δ contains two elements 0 (remain silent) and 1 (question the government). In Figure 3, each "initial" state is labeled by the set of actions safe for agent Luo Ji. Thus, for example, $S^{DDC}_{\text{Luo Ji}} = \{0, 1\}$. In our example, action ℓ is 0 (remain silent). We stipulate the existence of a default "safe" action in each state in order for each agent to be able to have at least one "safe" action. Existence of such safe action is important for the soundness of the Necessitation inference rule for modality H.

Informally, $(w, \delta, u) \in M$ means that the system can transition from state w to state u under action profile δ. For example, $(DDD, \delta_{100}, w_{\text{rev}}) \in M$ because if in state DDD Luo Ji questions the government (action 1) and Zhi Shi and Tui Li remain silent (action 0), then the game can transition into the revolution state w_{rev}. In general, a mechanism is a relation, not a function. Thus, transitions might be nondeterministic. If for some state $w \in W$ and some action profile $\delta \in \Delta^{\mathcal{A}}$ there is no state u such that $(w, \delta, u) \in M$, then we say that system terminates in state w under action profile δ. Note that unlike our examples, in general we do not distinguish "initial" and "outcome" states. We assume that after any transition the system might transition again using a different action profile.

Recall from the introduction that KS^w_a denotes the set of all actions that agent a *knows* to be *safe* for her in state w. In the censorship example, $KS^{DDC}_{\text{Luo Ji}} = KS^{DCC}_{\text{Luo Ji}} = \{0\}$.

Definition 3. *Let KS^w_a be the set of all actions $s \in \Delta$ such that $s \in S^{w'}_a$ for each state $w' \in W$ such that $w \sim_a w'$.*

Consider now an arbitrary partition σ of the set of all agents. If agents in the same partition communicate, they *might* learn that some additional actions are safe. By DS^w_σ we denote the set of all action profiles δ such that, for each agent $a \in \mathcal{A}$, the set of agents $[a]_\sigma$ *distributively knows* that action $\delta(a)$ is safe for a in state w. Informally, DS^w_σ is the set of all action profiles δ about which each agent $a \in \mathcal{A}$ *might* learn that action $\delta(a)$ is safe for her in state w if she communicates with the other agents in the set $[a]_\sigma$.

Definition 4. *DS^w_σ consists of all action profiles $\delta \in \Delta^{\mathcal{A}}$ such that for each agent $a \in \mathcal{A}$ and each state $w' \in W$, if $w \sim_b w'$ for each agent $b \in [a]_\sigma$, then $\delta(a) \in S^{w'}_a$.*

For example, $\delta_{100} \notin DS^{DDD}_{\tau_1}$, but $\delta_{100} \in DS^{DDD}_{\tau_2}$ because under partition τ_2 Luo Ji *might* learn that Zhi Shi is discontent. Thus, Luo Ji might learn that it is safe for him to question the government.

Lemma 1. *If $\sigma \preceq \tau$, then $DS^w_\sigma \subseteq DS^w_\tau$.* □

Next is the key definition of this paper. It gives formal semantics of modalities K, N, and H.

Definition 5. *For any state $w \in W$ and any formula $\varphi \in \Phi$, satisfiability relation $w \Vdash \varphi$ is defined as follows*

1. $w \Vdash p$ *if* $w \in \pi(p)$,
2. $w \Vdash \neg\varphi$ *if* $w \nVdash \varphi$,
3. $w \Vdash \varphi \to \psi$ *if* $w \nVdash \varphi$ *or* $w \Vdash \psi$,
4. $w \Vdash \mathsf{K}_a\varphi$ *if* $u \Vdash \varphi$ *for each* $u \in W$ *such that* $w \sim_a u$,
5. $w \Vdash \mathsf{N}^\sigma\varphi$ *when for each action profile* $\delta \in DS^w_\sigma$ *and each state* $u \in W$, *if* $(w, \delta, u) \in M$, *then* $u \Vdash \varphi$,
6. $w \Vdash \mathsf{H}^\sigma_a\varphi$ *when there is an action* $s \in KS^w_a$ *such that for all states* $w', u \in W$ *and each action profile* $\delta \in DS^{w'}_{\sigma/a}$, *if* $\delta(a) = s$, $w \sim_a w'$, *and* $(w', \delta, u) \in M$, *then* $u \Vdash \varphi$.

Informally, $\mathsf{N}^\sigma\varphi$ means that statement φ is true under any action profile δ in which each agent a takes an action $\delta(a)$ that she *might learn* is safe for her if she communicates with the other agents in set $[a]_\sigma$.

Formula $\mathsf{H}^\sigma_a\varphi$ denotes that, in presence of the information walls specified by partition σ, agent a knows a safe strategy to achieve φ. In the introduction we used the example

$$\mathsf{H}^\sigma_{\text{Al-Qaeda}}(\text{“flight 77 crashed into the Pentagon”}).$$

Note that FBI *might*, but is not *expected* to volunteer any information to Al-Qaeda. Thus, item 6 of the above definition requires the action s to be knowingly safe to agent a without communication with other agents. Although partition σ, see Figure 2, does not prevent Al-Qaeda and FBI from sharing information, we allow for a strategy of Al-Qaeda to rely on the fact that Al-Qaeda, while pursuing the strategy, will not volunteer any information to FBI. Hence, FBI would need to know that its actions are safe without relying on any information from Al-Qaeda. This is why item 6 in the above definition requires $\delta \in DS^{w'}_{\sigma/a}$ instead of $\delta \in DS^{w'}_\sigma$. Finally, state w' in item 6 is used to capture the fact that the desired strategy for agent a not only exists, but is also known by agent a. In other words, it is a know-how strategy.

Axioms

In additional to propositional tautologies in language Φ, our logical system contains the following axioms:[3]

1. Truth: $\mathsf{K}_a\varphi \to \varphi$,
2. Negative Introspection: $\neg\mathsf{K}_a\varphi \to \mathsf{K}_a\neg\mathsf{K}_a\varphi$,
3. Distributivity: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$ where $\Box \in \{\mathsf{K}_a, \mathsf{N}^\sigma\}$,
4. Monotonicity: $\mathsf{N}^\tau\varphi \to \mathsf{N}^\sigma\varphi$ where $\sigma \preceq \tau$, $\mathsf{H}^\tau_a\varphi \to \mathsf{H}^\sigma_a\varphi$ where $\sigma/a \preceq \tau/a$,
5. Strategic Introspection: $\mathsf{H}^\sigma_a\varphi \to \mathsf{K}_a\mathsf{H}^\sigma_a\varphi$,
6. Known Necessity: $\mathsf{K}_a\mathsf{N}^{\sigma/a}(\varphi \to \psi) \to (\mathsf{H}^\sigma_a\varphi \to \mathsf{H}^\sigma_a\psi)$.

The Truth, the Negative Introspection, and the Distributivity axioms are well-known modal properties. The Monotonicity axiom for modality N captures the fact that if something is unavoidably true under communication walls imposed by partition τ, then the same is also unavoidably true under any partition σ that has additional information walls. Similar property is true for modality H except that assumption $\sigma \preceq \tau$ is replaced with a weaker assumption $\sigma/a \preceq \tau/a$ because formal semantics of modality H^σ_a excludes communication between agent a and the other agents in class $[a]_\sigma$. The Strategic Introspection axiom states that if an agent has a know-how strategy, then she knows that she has a know-how strategy. The Known Necessity axiom states that if agent a knows that $\varphi \to \psi$ is unavoidable as long as agent a remains silent and she also knows how to achieve φ, then she knows how to achieve ψ. Formally, “agent a remains silent” is captured by using partition σ/a instead of partition σ.

We write $\vdash \varphi$, and say that φ is a *theorem* of our logical system, if formula φ is provable from the above axioms using the Modus Ponens, the three forms of the Necessitation, and the Monotonicity inference rules:

$$\frac{\varphi, \varphi \to \psi}{\psi} \qquad \frac{\varphi}{\mathsf{H}^\sigma_a\varphi} \qquad \frac{\varphi}{\mathsf{K}_a\varphi} \qquad \frac{\varphi}{\mathsf{N}^\sigma\varphi} \qquad \frac{\varphi \to \psi}{\mathsf{H}^\sigma_a\varphi \to \mathsf{H}^\sigma_a\psi}.$$

In addition to unary relation $\vdash \varphi$, we also consider binary relation $X \vdash \varphi$ which is true if formula φ is provable from the *theorems* of our logical system and the set of additional axioms X using only the Modus Ponens inference rule. Note that $\varnothing \vdash \varphi$ is equivalent to $\vdash \varphi$.

Soundness

In this section, we show the soundness of our logical system. Soundness of the Truth, the Negative Introspection, and the Distributivity axioms is standard. Below we prove soundness of each of the remaining axioms as a separate lemma.

Lemma 2. *If* $\sigma \preceq \tau$ *and* $w \Vdash \mathsf{N}^\tau\varphi$, *then* $w \Vdash \mathsf{N}^\sigma\varphi$.

Proof. Consider any action profile $\delta \in DS^w_\sigma$ and any state $u \in W$ such that $(w, \delta, u) \in M$. By item 5 of Definition 5, it suffices to show $u \Vdash \varphi$. Indeed, the assumption $\delta \in DS^w_\sigma$ and the assumption $\sigma \preceq \tau$ of the lemma imply $\delta \in DS^w_\tau$ by Lemma 1. Hence, $u \Vdash \varphi$ by the assumption $w \Vdash \mathsf{N}^\sigma\varphi$, item 5 of Definition 5, and the assumption $(w, \delta, u) \in M$. □

Lemma 3. *If* $\sigma/a \preceq \tau/a$ *and* $w \Vdash \mathsf{H}^\tau_a\varphi$, *then* $w \Vdash \mathsf{H}^\sigma_a\varphi$.

Proof. By the assumption $w \Vdash \mathsf{H}^\tau_a\varphi$ of the lemma and item 6 of Definition 5, there is an action $s \in KS^w_a$ such that for all states $w', u \in W$ and each action profile $\delta \in DS^{w'}_{\tau/a}$, if $\delta(a) = s$, $w \sim_a w'$, and $(w', \delta, u) \in M$, then $u \Vdash \varphi$.

Consider any states $w', u \in W$ and any action profile $\delta \in DS^{w'}_{\sigma/a}$ such that $\delta(a) = s$, $w \sim_a w'$, and $(w', \delta, u) \in M$. By item 6 of Definition 5, it suffices to show that $u \Vdash \varphi$.

Notice that the assumption $\delta \in DS^{w'}_{\sigma/a}$ and the assumption $\sigma/a \preceq \tau/a$ of the lemma imply $\delta \in DS^{w'}_{\tau/a}$ by Lemma 1. Therefore, $u \Vdash \varphi$ by the choice of action s using the assumptions $\delta(a) = s$, $w \sim_a w'$, and $(w', \delta, u) \in M$. □

Lemma 4. *If* $w \Vdash \mathsf{H}^\sigma_a\varphi$, *then* $w \Vdash \mathsf{K}_a\mathsf{H}^\sigma_a\varphi$.

Proof. Consider any state $v \in W$ such that $w \sim_a v$. By item 4 of Definition 5, it suffices to show that $v \Vdash \mathsf{H}^\sigma_a\varphi$.

By item 6 of Definition 5, the assumption $w \Vdash \mathsf{H}^\sigma_a\varphi$ of the lemma implies that there is an action $s \in KS^w_a$ such that for all states $w', u \in W$ and each action profile $\delta \in DS^{w'}_{\sigma/a}$, if $\delta(a) = s$, $w \sim_a w'$, and $(w', \delta, u) \in M$, then $u \Vdash \varphi$.

[3] Notations $\sigma \preceq \tau$ and σ/a have been introduced in the beginning of the previous section.

Then, by the assumption $w \sim_a v$, for all states $w', u \in W$ and each action profile $\delta \in DS^{w'}_{\sigma/a}$, if $\delta(a) = s$, $v \sim_a w'$, and $(w', \delta, u) \in M$, then $u \Vdash \varphi$. Therefore, $v \Vdash \mathsf{H}^{\sigma}_{a}\varphi$ by item 6 of Definition 5. □

Lemma 5. *If* $w \Vdash \mathsf{K}_a\mathsf{N}^{\sigma/a}(\varphi \to \psi)$ *and* $w \Vdash \mathsf{H}^{\sigma}_{a}\varphi$, *then* $w \Vdash \mathsf{H}^{\sigma}_{a}\psi$.

Proof. By item 6 of Definition 5, the assumption $w \Vdash \mathsf{H}^{\sigma}_{a}\varphi$ implies that there is an action $s \in KS^{w}_{a}$ such that for all states $w', u \in W$ and each action profile $\delta \in DS^{w'}_{\sigma/a}$, if $\delta(a) = s$, $w \sim_a w'$, and $(w', \delta, u) \in M$, then $u \Vdash \varphi$.

Consider any two states $w', u \in W$ and any action profile $\delta \in DS^{w'}_{\sigma/a}$ where $\delta(a) = s$, $w \sim_a w'$, and $(w', \delta, u) \in M$. By item 6 of Definition 5, it suffices to show that $u \Vdash \psi$. Indeed, by item 4 of Definition 5, the assumption $w \sim_a w'$ and the assumption of the lemma $w \Vdash \mathsf{K}_a\mathsf{N}^{\sigma/a}(\varphi \to \psi)$ imply that $w' \Vdash \mathsf{N}^{\sigma/a}(\varphi \to \psi)$. Hence, $u \Vdash \varphi \to \psi$ by item 5 of Definition 5 and the assumptions $\delta \in DS^{w'}_{\sigma/a}$ and $(w, \delta, u) \in M$.

At the same time, $u \Vdash \varphi$ by the choice of action s and because $\delta(a) = s$, $w \sim_a w'$, and $(w', \delta, u) \in M$. Therefore, $u \Vdash \psi$ by item 3 of Definition 5. □

Key Ideas behind the Proof of the Completeness

Distributed Key Generation

The standard proof of completeness for the multiagent version of epistemic logic S5 defines the states of the canonical model as maximal consistent sets of formulae. Two such states are a-indistinguishable if they contain the same K_a-formulae. This construction does not work in our case because we allow formulae that simultaneously use modality H^{σ}_{a} for different partitions σ. Indeed, recall agents Luo Ji, Zhi Shi, and Tui Li from one of the introductory examples. Consider any maximal consistent set of formulae w that contains exactly the same $\mathsf{K}_{\text{Zhi Shi}}$- and $\mathsf{K}_{\text{Luo Ji}}$-formulae. In other words, for any formula $\varphi \in \Phi$,

$$\mathsf{K}_{\text{Luo Ji}}\varphi \in w \quad \text{iff} \quad \mathsf{K}_{\text{Zhi Shi}}\varphi \in w.$$

If the indistinguishability relation is defined as in the standard construction, the equivalence classes of state w with respect to relation $\sim_{\text{Luo Ji}}$ and relation $\sim_{\text{Zhi Shi}}$ would be the same. Thus, if the canonical model is defined in the standard way, then agents Luo Ji and Zhi Shi will know exactly the same in state w.

Next, suppose that set w contains formulae $\mathsf{H}^{\tau_1}_{\text{Tui Li}}\varphi$ and $\neg\mathsf{H}^{\tau_2}_{\text{Tui Li}}\varphi$ for some formula $\varphi \in \Phi$, were partitions τ_1 and τ_2 are specified in Figure 4. For the reader's convenience, we reproduce this figure here as Figure 5. The key step in the standard proof of the completeness is the "truth" (or "induction") lemma that states that a formula belongs to set w if and only if it is satisfied in state w. In our case, this lemma would imply that $w \Vdash \mathsf{H}^{\tau_1}_{\text{Tui Li}}\varphi$ and $w \Vdash \neg\mathsf{H}^{\tau_2}_{\text{Tui Li}}\varphi$.

Statements $w \Vdash \mathsf{H}^{\tau_1}_{\text{Tui Li}}\varphi$ and $w \Vdash \neg\mathsf{H}^{\tau_2}_{\text{Tui Li}}\varphi$ mean that agent Tui Li has a know-how strategy to achieve φ when the wall between Luo Ji and Zhi Shi is present and does not have such a strategy otherwise. Informally, this should happen because, in the absence of the wall, information can freely travel between Luo Ji and Zhi Shi and thus, they both have larger sets of knowingly safe actions. If they use strategies from these larger sets, then Tui Li's strategy might no longer work. However, as we have seen above, if the standard construction is used to build the canonical model, then Luo Ji and Zhi Shi have exactly the same knowledge and, thus, there is absolutely nothing new that they can learn by sharing information with each other!

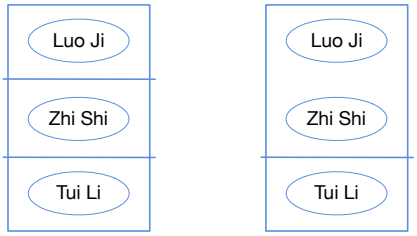

Figure 5: Partitions τ_1 (left) and τ_2 (right).

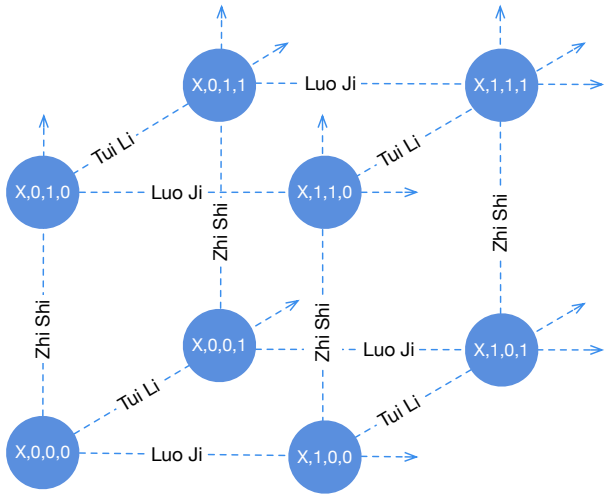

Figure 6: States of the canonical model for a single maximal consistent set X.

To overcome this issue, we need Luo Ji and Zhi Shi to possess some additional knowledge that they do not have under the standard canonical model construction. We add this knowledge to our canonical model using the *distributed key generation*. This is a cryptographic technique consisting in independent generation of random keys by several agents (Pedersen 1991). In our running example, each state of the canonical model will be a quadruple (X, l, z, t), where X is a maximal consistent set of formulae and l, z, t are integer "keys" of Luo Ji, Zhi Shi, and Tui Li respectively. We assume that each agent knows her own key, but not the keys of the other agents. For the same maximal consistent set X, such states and the three indistinguishability relations between them are partially depicted in Figure 6. The complete infinite set of states consists of all quadruples (X, l, z, t) for

all possible maximal consistent set X and all integer values l, z, and t. Incorporation of the distributed key generation into epistemic model construction is a new idea that we introduce in this paper.

Harmony

As mentioned earlier, the proofs of completeness usually use a "truth" or "induction" lemma that states that $\varphi \in w$ if and only if $w \Vdash \varphi$ for any formula φ and any state w. It claims that $\varphi \in X_w$ iff $w \Vdash \varphi$, where X_w is the first component of state w, as discussed in the previous section.

Consider now the case when formula φ has the form $\mathsf{K}_a\psi$. If $\mathsf{K}_a\psi \notin X_w$, then, by item 4 of Definition 5, the canonical model construction must guarantee that $w \nVdash \mathsf{K}_a\psi$. As usual, we achieve this by using Lindenbaum's lemma to construct a new state u such that $w \sim_a u$ and $u \nVdash \psi$.

The situation is more complicated if formula φ has the form $\mathsf{H}_a^\sigma\psi$. In this case, the canonical model must contain *two* different states, w' and u, satisfying conditions stated in item 6 of Definition 5. An important step in creating these two states is the construction of the corresponding maximal consistent sets $X_{w'}$ and X_u. It turns out that these two sets cannot be created consecutively.

Naumov and Tao (2018c; 2018a) proposed a technique called *harmony* for simultaneous construction of two maximal consistent sets. Their technique cannot be directly applied in our setting because the original harmony was not designed to deal with information walls in the set of agents. In this paper we propose a variation of their technique that we call σ-harmony.

The technique consists of identifying a certain invariant condition on a pair of sets of formulae, proving that an "initial" pair of sets satisfies this condition, and showing that the sets could be expanded while preserving the invariant. We call the invariant condition σ-harmony, just like the technique itself. The expansion step is repeated infinitely many times to achieve another condition, that we call *complete* σ-harmony. As a final step, Lindenbaum's lemma is used to "top-off" the two sets in complete σ-harmony to maximal consistent sets.

Due to space constraints, the full proof of the following strong completeness theorem is in the full version of this paper.

Theorem 1. *If $Y \nvdash \varphi$, then there is a state w of a game such that $w \Vdash \chi$ for each formula $\chi \in Y$ and $w \nVdash \varphi$.*

Conclusion

The contribution of this paper is two-fold. First, we introduced a new class of strategies that rely on presence of "information walls" between players. Second, we proposed a sound and complete modal logic that describes the properties of such strategies in games with imperfect information.

Perhaps the most natural question about this work is if the current results could be generalized to group knowledge and coalition know-how strategies. One of the challenges in this direction is finding an intuitively acceptable interpretation of group knowledge in the presence of information walls. Is it sensible to reason about a coalition distributively knowing a strategy if the coalition members are on different sides of a wall and explicitly banned from communicating with each other? One might consider only coalitions C located in the same set of a partition, but this makes the syntax confusing given that we study modalities H_C^σ for different partitions σ.

We think that a more interesting direction is to study one-way information walls that only prevent diffusion of the information in one of two directions. In real-world scenarios, for example, certain group of people might be banned from spreading information to outsiders, but not from listening to them.

Another possible extension of this work is to consider walls that do not block information diffusion completely, but impose cost on it. In such a setting, for instance, one can study modality $\mathsf{H}_a^m\varphi$ that stands for "agent a knows a strategy to achieve φ as long as the total cost of communication by all agents is no more than m. In a similar way, one can introduce degrees of "safeness".

Finally, another interesting direction for future research is studying group actions that require *common knowledge* of safeness. For instance, in the famous example with two generals, the generals are not able start a joint attack on a common enemy because they cannot establish common knowledge of the time to attack. The two general setting is very similar to the setting of this paper if the notion of distributively know safe action from Definition 4 is replaced with commonly known safe action. In this modified setting we could, for example, express the fact that the common enemy has a strategy to win the battle with the two generals because they will never be able to start a coordinated counterattack.

References

Ågotnes, T.; and Alechina, N. 2019. Coalition Logic with Individual, Distributed and Common Knowledge. *Journal of Logic and Computation*, 29: 1041–1069.

Ågotnes, T.; Balbiani, P.; van Ditmarsch, H.; and Seban, P. 2010. Group announcement logic. *Journal of Applied Logic*, 8(1): 62 – 81.

Ågotnes, T.; van der Hoek, W.; and Wooldridge, M. 2009. Reasoning about coalitional games. *Artificial Intelligence*, 173(1): 45 – 79.

Alur, R.; Henzinger, T. A.; and Kupferman, O. 2002. Alternating-time temporal logic. *Journal of the ACM*, 49(5): 672–713.

Aminof, B.; Malvone, V.; Murano, A.; and Rubin, S. 2018. Graded modalities in Strategy Logic. *Inf. Comput.*, 261(Part): 634–649.

Aminof, B.; Murano, A.; Rubin, S.; and Zuleger, F. 2016. Prompt Alternating-Time Epistemic Logics. *KR*, 16: 258–267.

Belardinelli, F. 2014. Reasoning about Knowledge and Strategies: Epistemic Strategy Logic. In *Proceedings 2nd International Workshop on Strategic Reasoning, SR 2014, Grenoble, France, April 5-6, 2014*, volume 146 of *EPTCS*, 27–33.

Belnap, N.; and Perloff, M. 1990. Seeing to it that: A canonical form for agentives. In *Knowledge representation and defeasible reasoning*, 167–190. Springer.

Berthon, R.; Maubert, B.; and Murano, A. 2017. Decidability results for ATL* with imperfect information and perfect recall. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 1250–1258. International Foundation for Autonomous Agents and Multiagent Systems.

Berthon, R.; Maubert, B.; Murano, A.; Rubin, S.; and Vardi, M. Y. 2017. Strategy logic with imperfect information. In *Logic in Computer Science (LICS), 2017 32nd Annual ACM/IEEE Symposium on*, 1–12. IEEE.

Borgo, S. 2007. Coalitions in Action Logic. In *20th International Joint Conference on Artificial Intelligence*, 1822–1827.

Broersen, J.; Herzig, A.; and Troquard, N. 2007. A normal simulation of coalition logic and an epistemic extension. In *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, 92–101. ACM.

Chatterjee, K.; Henzinger, T. A.; and Piterman, N. 2010. Strategy logic. *Information and Computation*, 208(6): 677–693.

Donders, M. S.; More, S. M.; and Naumov, P. 2011. Information Flow on Directed Acyclic Graphs. In Beklemishev, L. D.; and de Queiroz, R., eds., *WoLLIC*, volume 6642 of *Lecture Notes in Computer Science*, 95–109. Springer. ISBN 978-3-642-20919-2.

Fervari, R.; Herzig, A.; Li, Y.; and Wang, Y. 2017. Strategically knowing how. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 1031–1038.

Goranko, V. 2001. Coalition games and alternating temporal logics. In *Proceedings of the 8th conference on Theoretical aspects of rationality and knowledge*, 259–272. Morgan Kaufmann Publishers Inc.

Goranko, V.; and van Drimmelen, G. 2006. Complete axiomatization and decidability of Alternating-time temporal logic. *Theoretical Computer Science*, 353(1): 93 – 117.

Grewe, B. A. 2004. Legal Barriers to Information Sharing: The Erection of a Wall Between Intelligence and Law Enforcement Investigations. Technical report, Commission on Terrorist Attacks Upon the United States.

Hayes, J. 2017. Private Sector Workers Lack Pay Transparency: Pay Secrecy May Reduce Women's Bargaining Power and Contribute to Gender Wage Gap. Technical Report IWPR#Q068, The Institute for Women's Policy Research.

Horty, J.; and Pacuit, E. 2017. Action types in STIT semantics. *The Review of Symbolic Logic*, 10(4): 617–637.

Horty, J. F. 2001. *Agency and deontic logic*. Oxford University Press.

Horty, J. F.; and Belnap, N. 1995. The deliberative STIT: A study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, 24(6): 583–644.

Kane, J.; and Naumov, P. 2013. Epistemic Logic for Communication Chains. In *14th conference on Theoretical Aspects of Rationality and Knowledge (TARK '13), January 2013, Chennai, India*, 131–137.

Kean, T. 2004. *The 9/11 commission report: Final report of the national commission on terrorist attacks upon the United States*. Government Printing Office.

Mogavero, F.; Murano, A.; Perelli, G.; and Vardi, M. Y. 2014. Reasoning about strategies: On the model-checking problem. *ACM Transactions on Computational Logic (TOCL)*, 15(4): 34.

More, S. M.; and Naumov, P. 2011a. Hypergraphs of multiparty secrets. *Ann. Math. Artif. Intell.*, 62(1-2): 79–101.

More, S. M.; and Naumov, P. 2011b. Logic of secrets in collaboration networks. *Ann. Pure Appl. Logic*, 162(12): 959–969.

More, S. M.; and Naumov, P. 2012. Calculus of Cooperation and Game-based Reasoning About Protocol Privacy. *ACM Trans. Comput. Logic*, 13(3): 22:1–22:21.

Naumov, P.; and Tao, J. 2017a. Coalition Power in Epistemic Transition Systems. In *Proceedings of the 2017 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 723–731.

Naumov, P.; and Tao, J. 2018a. Second-Order Know-How Strategies. In *Proceedings of the 2018 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 390–398.

Naumov, P.; and Tao, J. 2018b. Strategic Coalitions with Perfect Recall. In *Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence*.

Naumov, P.; and Tao, J. 2018c. Together We Know How to Achieve: An Epistemic Logic of Know-How. *Artificial Intelligence*, 262: 279 – 300.

Naumov, P.; and Tao, J. 2019. Knowing-how under uncertainty. *Artificial Intelligence*, 276: 41 – 56.

Naumov, P. G.; and Tao, J. 2017b. Knowledge in communication networks. *Journal of Logic and Computation*, 27(4): 1189–1224.

Olkhovikov, G. K.; and Wansing, H. 2019. Inference as doxastic agency. Part I: The basics of justification STIT logic. *Studia Logica*, 107(1): 167–194.

Pacuit, E.; and Parikh, R. 2004. The logic of communication graphs. In *International Workshop on Declarative Agent Languages and Technologies*, 256–269. Springer.

Pacuit, E.; and Parikh, R. 2007. Reasoning about communication graphs. *Interactive Logic*, 1: 135–157.

Pauly, M. 2001. *Logic for Social Software*. Ph.D. thesis, Institute for Logic, Language, and Computation.

Pauly, M. 2002. A Modal Logic for Coalitional Power in Games. *Journal of Logic and Computation*, 12(1): 149–166.

Pedersen, T. P. 1991. Non-interactive and information-theoretic secure verifiable secret sharing. In *Annual international cryptology conference*, 129–140. Springer.

Sauro, L.; Gerbrandy, J.; van der Hoek, W.; and Wooldridge, M. 2006. Reasoning About Action and Cooperation. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '06, 185–192. New York, NY, USA: ACM. ISBN 1-59593-303-4.

Sutherland, D. 1986. A Model of Information. In *Proceedings of Ninth National Computer Security Conference*, 175–183.

van der Hoek, W.; and Wooldridge, M. 2003. Cooperation, knowledge, and time: Alternating-time temporal epistemic logic and its applications. *Studia Logica*, 75(1): 125–157.

van der Hoek, W.; and Wooldridge, M. 2005. On the logic of cooperation and propositional control. *Artificial Intelligence*, 164(1): 81 – 119.

Wang, Y. 2015. A logic of knowing how. In *Logic, Rationality, and Interaction*, 392–405. Springer.

Wang, Y. 2018. A logic of goal-directed knowing how. *Synthese*, 195(10): 4419–4439.