# Second-Order Know-How Strategies

Pavel Naumov
Vassar College
Poughkeepsie, NY
pnaumov@vassar.edu

Jia Tao
Lafayette College
Easton, PA
taoj@lafayette.edu

## ABSTRACT

The fact that a coalition has a strategy does not mean that the coalition knows what the strategy is. If the coalition knows the strategy, then such a strategy is called a know-how strategy of the coalition. The paper proposes the notion of a second-order know-how strategy for the case when one coalition knows what the strategy of another coalition is. The main technical result is a sound and complete logical system describing the interplay between the distributed knowledge modality and the second-order coalition know-how modality.

## KEYWORDS

coalition power, knowledge, formal epistemology, strategy, game theory, transition system, multiagent system, modal logic, axiomatization, completeness

## 1 INTRODUCTION

In this paper we study the interplay between coalition strategies and the distributed knowledge in multiagent systems.

### 1.1 Coalition Strategies

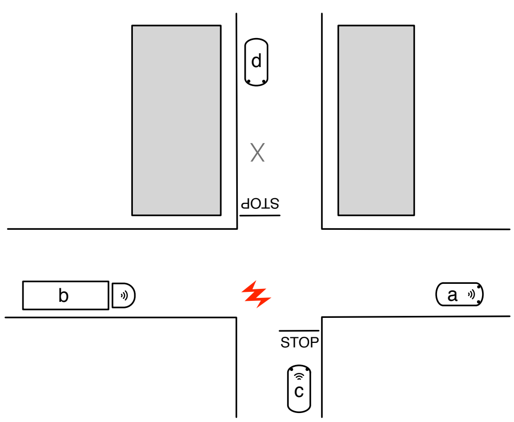


**Figure 1: State of Traffic $w_1$.**

Consider the traffic situation depicted in Figure 1, where a regular vehicle $d$ and three self-driving vehicles $a$, $b$, and $c$ approach an intersection. There are stop signs at the intersection facing cars $c$ and $d$. According to the traffic rules, these two cars must slow down, stop, and yield to truck $b$.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018,Stockholm, Sweden. 

Suppose that the driver of car $d$ does not notice the sign and, as a result, this car is approaching the intersection with a constant speed. If neither of the vehicles changes its behavior, car $d$ will hit the side of the rear half of truck $b$, at the location marked with a red zigzag shape on the figure. Truck $b$ can potentially slow down, but then car $d$ will hit the side of truck $b$ in the front half instead of the rear half. Thus, to avoid being struck by car $d$, truck $b$ has to accelerate and pass the intersection before car $d$ does. We assume that in this case if car $a$ maintains the same speed, then there will be a rear-end collision between truck $b$ and car $a$. Hence, to avoid any collision, not only must truck $b$ accelerate, but car $a$ must accelerate as well. In other words, to prevent a collision, vehicles $a$ and $b$ must engage in a strategic cooperation. We say that coalition $\{a, b\}$ has a *strategy* to prevent a collision.

The traffic situation is further complicated by two buildings, shown in Figure 1 as grey rectangles. The buildings prevent car $a$ and truck $b$ from seeing car $d$. Although coalition $\{a, b\}$ has a strategy to avoid collision, it does not know what this strategy is, nor does it know that such a strategy exists. However, self-driving car $c$ can observe that car $d$ is not slowing down, and it can make coalition $\{a, b\}$ aware of the presence of car $d$ as well as its speed and location. With the information shared by car $c$, not only will coalition $\{a, b\}$ have a strategy to avoid a collision, but it also will know what this strategy is.

In general, the following cases might take place: (i) a coalition does not have a strategy; (ii) a coalition has a strategy, but it does not know that it has a strategy; (iii) a coalition knows that it has a strategy, but it does not know what the strategy is; (iv) a coalition knows that it has a strategy and it knows what this strategy is. In the last case, we say that the coalition has a *know-how* strategy. Know-how strategies were studied before under different names. While Jamroga and Ågotnes talked about “knowledge to identify and execute a strategy” [5], Jamroga and van der Hoek discussed “difference between an agent knowing that he has a suitable strategy and knowing the strategy itself” [6]. Van Benthem called such strategies “uniform” [16]. Broersen talked about “knowingly doing” [2], while Broersen, Herzig, and Troquard discussed modality “know they can do” [3].

In our example, coalition $\{a, b\}$ has a know-how strategy to avoid a collision after car $c$ shares the traffic information with the coalition. In other words, it is not coalition $\{a, b\}$ but the single-element coalition $\{c\}$ that knows what is the strategy of coalition $\{a, b\}$. We refer to such strategies as *second-order* know-how strategies by analogy with the commonly used term *second-order knowledge* [7].

Second-order know-how manifests itself in many settings. For example, a teacher might know how a student can succeed or a group of campaign advisers might know how a political party can win the elections.

## 1.2 Epistemic Transition Systems

We use the notion of an *epistemic transition system* to formalize the concept of a second-order know-how strategy. A fragment of an epistemic transition system corresponding to the traffic situation described above is depicted on the diagram in Figure 2. In particular, the traffic situation depicted in Figure 1 is represented by state $w_1$ on this diagram. The arrows on the diagram correspond to possible transitions of the system.

For the sake of simplicity, we assume that transitions of this system only depend on the actions of agents $a$ and $b$. Moreover, each of agents $a$ and $b$ is assumed to have just three strategies: to slow down (–), to maintain current speed (0), and to accelerate (+). In Figure 2, transitions are labeled by the strategies of agents $a$ and $b$ that accomplish the transition.

If a transition is labeled with strategy profile $(x, y)$, then $x$ represents the strategy of car $a$ and $y$ represents the strategy of truck $b$. Although there are nine possible transitions from state $w_1$, corresponding to nine possible strategy profiles $(x, y)$, the fragment of this system (depicted in Figure 2) shows only four such transitions leading to states $w_4, w_5, w_6$, and $w_7$.

As we discussed earlier, vehicles $a$ and $b$ can use coalition strategy $(+,+)$ to avoid a collision. However, they *do not know* that this coalition strategy would prevent a collision because they do not even know the presence of vehicle $d$, let alone its location and speed. To show this formally, consider a hypothetical state $w_2$ in Figure 2. In this state, vehicle $d$ is currently at the spot marked by symbol X on Figure 1. In this state, car $d$ is closer to the intersection than it is in state $w_1$. A simultaneous acceleration of vehicles $a$ and $b$ (coalition strategy $(+,+)$) would not prevent a collision because car $d$ would hit the side of truck $b$ in the rear half. Instead, in state $w_2$, coalition $\{a, b\}$ can use, for example, strategy $(0, -)$ to avoid a collision. Under this strategy, car $a$ maintains the current speed and truck $b$ slows down. Since vehicles $a$ and $b$ cannot see vehicle $d$, they cannot distinguish states $w_1$ and $w_2$. We define a coalition *know-how strategy* at state $w_1$ as a strategy that would succeed in all states indistinguishable from state $w_1$ by the coalition. Thus, the transition system whose fragment is depicted on Figure 2 does not have a know-how strategy for coalition $\{a, b\}$ to avoid a collision in state $w_1$. States $w_1$ and $w_2$ are not the only indistinguishable states in this system. For example, state $w_3$ in the same figure, where car $d$ is not present at the scene, is also indistinguishable from states $w_1$ and $w_2$ to coalition $\{a, b\}$.

## 1.3 Second-Order Know-How Modality

Recall that self-driving car $c$ can observe that car $d$ is not slowing down. Thus, car $c$ can distinguish states $w_1$ from states $w_2$ and $w_3$ of this system. The system might have other states indistinguishable to car $c$ from state $w_1$. These states, for example, could differ by traffic situations on nearby streets. However, in all these states, coalition $\{a, b\}$ can use strategy $(+,+)$ to avoid a collision. Hence, we say that agent $c$ *knows* how coalition $\{a, b\}$ can avoid a collision. We denote this fact by $w_1 \Vdash \mathsf{H}^{\{a,b\}}_{\{c\}}$(“avoid a collision”). In general, we write $w \Vdash \mathsf{H}^D_C\varphi$ if coalition $C$ has distributed knowledge of how coalition $D$ can achieve outcome $\varphi$ from state $w$. We call modality $\mathsf{H}$ the *second-order know-how modality*. Although in our example coalitions $D = \{a, b\}$ and $C = \{c\}$ are disjoint, we do allow these coalitions to have common elements. Modality $\mathsf{H}^C_C$ expresses the existence of a know-how strategy of coalition $C$ known to the coalition itself. Thus, it expresses an existence of a *first-order* know-how strategy of coalition $C$.

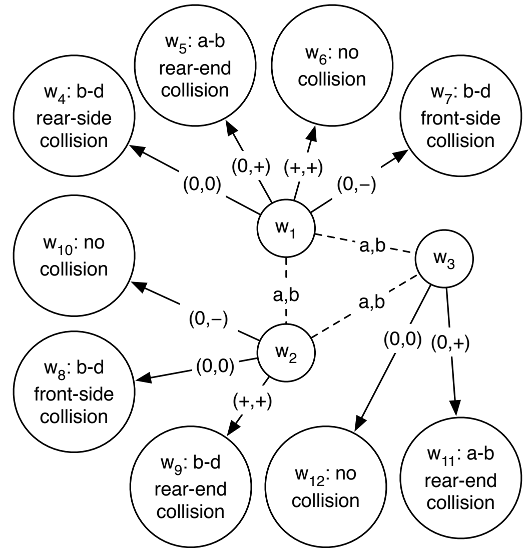


**Figure 2: A fragment of an epistemic transition system.**

Properties of first-order know-how modalities and their interplay with different forms of the knowledge modality have been studied before. Ågotnes and Alechina [1] proposed a complete axiomatization of an interplay between single-agent knowledge and first-order coalition know-how modalities to achieve a goal in one step. A modal logic that combines the distributed knowledge modality with the first-order coalition know-how modality to maintain a goal was axiomatized by Naumov and Tao [8]. 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 is developed by Fervari, Herzig, Li, and Wang [4]. A trimodal logical system that describes an interplay between the (not know-how) coalition strategic modality, the first-order coalition know-how modality, and the distributed knowledge modality was developed by Naumov and Tao [10]. They also proposed a logical system that combines the first-order coalition know-how modality with the distributed knowledge modality in the perfect recall setting [9, 12]. Wang proposed a complete axiomatization of “knowing how” as a binary modality [17, 18], but his logical system does not include the knowledge modality.

The main goal of this paper is to describe the interplay between the *second-order* know-how modality $\mathsf{H}$ and the distributed knowledge modality $\mathsf{K}$. In other words, we axiomatize all properties in the bimodal language that are true in all states of all epistemic transition systems. In addition to the distributed version of S5 axioms for modality $\mathsf{K}$, our logical system contains the Cooperation axiom, introduced by Marc Pauly [13, 14] for strategies in general,

$$\mathsf{H}^{D_1}_{C_1}(\varphi \to \psi) \to (\mathsf{H}^{D_2}_{C_2}\varphi \to \mathsf{H}^{D_1\cup D_2}_{C_1\cup C_2}\psi), \tag{1}$$

where $D_1 \cap D_2 = \varnothing$. Informally, this axiom states that second-order know-how strategies of two disjoint coalitions can be combined to form a single second-order know-how strategy to achieve a common goal. The system also has the Strategic Introspection axiom:

$$\mathsf{H}_C^D\varphi \to \mathsf{K}_C\mathsf{H}_C^D\varphi, \tag{2}$$

which states that if coalition $C$ knows how coalition $D$ can achieve the goal, then coalition $C$ knows that it knows how coalition $D$ can achieve this. A version of this axiom for the first-order know-how appeared in [1]. In addition, our logical system contains the Empty coalition axiom which appeared first in [10]:

$$\mathsf{K}_\varnothing\varphi \to \mathsf{H}_\varnothing^\varnothing\varphi. \tag{3}$$

This axiom says that if a statement is known to the empty coalition, then the empty coalition has a first-order know-how strategy to achieve it. The axiom is true because the empty coalition can know only statements that are true in each state of the given epistemic transition system. The final and perhaps the most interesting axiom of our logical system is the Knowledge of Unavoidability axiom:

$$\mathsf{K}_A\mathsf{H}_B^\varnothing\varphi \to \mathsf{H}_A^\varnothing\varphi. \tag{4}$$

Formula $\mathsf{H}_B^\varnothing\varphi$ means that coalition $B$ knows that $\varphi$ will be achieved no matter how agents act. Thus, coalition $B$ knows that $\varphi$ is unavoidable. The axiom states that if coalition $A$ knows that coalition $B$ knows that $\varphi$ is unavoidable, then coalition $A$ also knows that $\varphi$ is unavoidable. To the best of our knowledge, this axiom does not appear in the existing literature. The main technical results of this paper are the soundness and the completeness of the logical system describing the interplay between modalities K and H. The system extends epistemic logic S5 with distributed knowledge by axioms (1), (2), (3), and (4).

## 1.4 Harmony

Proving the completeness theorem for the interplay between knowledge and second-order know-how modalites is significantly more difficult than proving corresponding completeness theorems for bimodal logical systems describing the interplay between knowledge and first-order know-how modalities [1, 4, 8, 9, 12]. In the proof we use notions of harmony and complete harmony (see Definition 6.2 and Definition 6.6). The proof technique based on harmony has been originally developed by Naumov and Tao [10] for the *trimodal* logical system that describes the interplay between a distributed knowledge modality, a (not know-how) strategic modality, and a first-order know-how modality. We have modified the definitions of harmony and complete harmony for this technique to work in the current setting. See Section 6.1 and Section 6.2 for details.

## 1.5 Paper Outline

This paper is organized as follows. In Section 2 we introduce formal syntax and semantics of the logical system. In Section 3 we list its axioms and inference rules. Section 4 provides examples of formal proofs. The proofs of the soundness and the completeness are presented in Sections 5 and 6. Section 7 concludes the paper.

# 2 SYNTAX AND SEMANTICS

Throughout the paper we fix a countable set of propositional variables and a countable set of agents $\mathcal{A}$. A *coalition* is an arbitrary subset of $\mathcal{A}$. A *strategy profile* of a coalition $C$ is an arbitrary function that assigns a value from some domain $\Delta$ to each agent in coalition $C$. We denote the set of all such strategy profiles by $C^\Delta$. A *complete* strategy profile is a strategy profile of the coalition $\mathcal{A}$.

Following the convention in game theory, we consider a strategy profile as a (possibly infinite) tuple of values from $\Delta$ indexed by set $C$. If $\mathbf{s} \in C^\Delta$ and $a \in C$, then by $(\mathbf{s})_a$ we denote the component of this tuple corresponding to the index value $a$, which technically is the value of function $\mathbf{s}$ on the argument $a$.

*Definition 2.1.* A tuple $(W, \{\sim_a\}_{a\in\mathcal{A}}, \Delta, M, \pi)$ is an (epistemic) transition system, if

1. $W$ is a set (of epistemic states),
2. $\sim_a$ is an indistinguishability equivalence relation on set $W$ for each $a \in \mathcal{A}$,
3. $\Delta$ is a nonempty set, called the domain of actions,
4. $M \subseteq W \times \Delta^{\mathcal{A}} \times W$ is an aggregation mechanism,
5. $\pi$ is a function from propositional variables to subsets of $W$.

A fragment of a transition system is depicted in Figure 2. In this example, set $W$ consists of states such as $w_1, w_2$, and $w_3$. Relation $\sim_a$ is denoted by dashed lines. The domain of actions $\Delta$ is set $\{-, 0, +\}$. For each state, mechanism $M$ specifies the next state based on the actions of individual agents. The mechanism is captured by the directed edges between the states, labeled with strategy profiles.

There are two important things to note about the aggregation mechanism. First, the mechanism might be non-deterministic and thus we formally define it as a ternary relation between the current state, the strategy profile, and the next state. Second, the next state might not exist. If the next state does not exist for the selected strategy profile, then the transition system terminates.

*Definition 2.2.* 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 2.3. *For each coalition $C$, relation $\sim_C$ is an equivalence relation on the set of epistemic states $W$.* □

*Definition 2.4.* For any strategy profiles $\mathbf{s}_1$ and $\mathbf{s}_2$ of coalitions $C_1$ and $C_2$ respectively and any coalition $C \subseteq C_1 \cap C_2$, let $\mathbf{s}_1 =_C \mathbf{s}_2$ if $(\mathbf{s}_1)_a = (\mathbf{s}_2)_a$ for each $a \in C$.

LEMMA 2.5. *For any coalition $C$, relation $=_C$ is an equivalence relation on the set of all strategy profiles of coalitions containing coalition $C$.* □

*Definition 2.6.* Let $\Phi$ be the minimal set of formulae such that

1. $p \in \Phi$ for each propositional variable $p$,
2. $\neg\varphi, \varphi \to \psi \in \Phi$ for all formulae $\varphi, \psi \in \Phi$,
3. $\mathsf{K}_C\varphi, \mathsf{H}_C^D\varphi \in \Phi$ for each coalition $C$, each finite coalition $D$, and each formula $\varphi \in \Phi$.

It is crucial in the proof of completeness that superscript $D$ of modality $\mathsf{H}_C^D$ is finite. For the sake of generality, we allow subscript $C$ to be infinite. We assume that the constant $\top$ and the conjunction $\wedge$ are defined through connectives $\neg$ and $\to$ in the standard way. Furthermore, for any finite set of formulae $X$, by $\wedge X$ we mean the conjunction of all formulae in set $X$. In particular, $\wedge\varnothing$ is formula $\top$.

The next definition is the key definition of this paper. Its part (5) specifies the semantics of the second-order know-how modality $\mathsf{H}_C^D$. Informally, $w \Vdash \mathsf{H}_C^D\varphi$ means that there is a strategy of the coalition $D$ that can be used to achieve $\varphi$ from every state indistinguishable from state $w$ by the coalition $C$.

*Definition 2.7.* For any epistemic state $w \in W$ of a transition system $(W, \{\sim_a\}_{a\in\mathcal{A}}, V, M, \pi)$ and any formula $\varphi \in \Phi$, let relation $w \Vdash \varphi$ be defined as follows:

(1) $w \Vdash p$ if $w \in \pi(p)$, where $p$ is a propositional variable,
(2) $w \Vdash \neg\varphi$ if $w \nVdash \varphi$,
(3) $w \Vdash \varphi \to \psi$ if $w \nVdash \varphi$ or $w \Vdash \psi$,
(4) $w \Vdash \mathsf{K}_C\varphi$ if $w' \Vdash \varphi$ for each $w' \in W$ such that $w \sim_C w'$,
(5) $w \Vdash \mathsf{H}^D_C\varphi$ if there is a strategy profile $\mathbf{s} \in V^D$ such that for any two states $w', u \in W$ and any complete strategy profile $\mathbf{s}'$, if $w \sim_C w'$, $\mathbf{s} =_D \mathbf{s}'$, and $(w', \mathbf{s}', u) \in M$, then $u \Vdash \varphi$.

## 3 AXIOMS

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

(1) Truth: $\mathsf{K}_C\varphi \to \varphi$,
(2) Negative Introspection: $\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\varphi$,
(3) Distributivity: $\mathsf{K}_C(\varphi \to \psi) \to (\mathsf{K}_C\varphi \to \mathsf{K}_C\psi)$,
(4) Monotonicity: $\mathsf{K}_C\varphi \to \mathsf{K}_D\varphi$, if $C \subseteq D$,
(5) Cooperation: $\mathsf{H}^{D_1}_{C_1}(\varphi \to \psi) \to (\mathsf{H}^{D_2}_{C_2}\varphi \to \mathsf{H}^{D_1\cup D_2}_{C_1\cup C_2}\psi)$, where $D_1 \cap D_2 = \varnothing$.
(6) Strategic Introspection: $\mathsf{H}^D_C\varphi \to \mathsf{K}_C\mathsf{H}^D_C\varphi$,
(7) Empty Coalition: $\mathsf{K}_\varnothing\varphi \to \mathsf{H}^\varnothing_\varnothing\varphi$.
(8) Knowledge of Unavoidability: $\mathsf{K}_A\mathsf{H}^\varnothing_B\varphi \to \mathsf{H}^\varnothing_A\varphi$.

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

$$\frac{\varphi}{\mathsf{K}_C\varphi} \qquad \frac{\varphi}{\mathsf{H}^D_C\varphi} \qquad \frac{\varphi, \quad \varphi \to \psi}{\psi}.$$

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

## 4 EXAMPLES OF DERIVATIONS

We show the soundness of the above logical system in Section 5. Below, we provide some examples of formal proofs in this system. These results are used later in the proof of the completeness.

Lemma 4.1. *$\vdash \mathsf{H}^{D_1}_{C_1}\varphi \to \mathsf{H}^{D_2}_{C_2}\varphi$ where $C_1 \subseteq C_2$ and $D_1 \subseteq D_2$.*

Proof. Statement $\varphi \to \varphi$ is a tautology. Thus, by the Strategic Necessitation inference rule, $\vdash \mathsf{H}^{D_2\setminus D_1}_{C_2\setminus C_1}(\varphi \to \varphi)$. Next, by the Cooperation axiom and due to $C_1 \subseteq C_2$ and $D_1 \subseteq D_2$,

$$\vdash \mathsf{H}^{D_2\setminus D_1}_{C_2\setminus C_1}(\varphi \to \varphi) \to (\mathsf{H}^{D_1}_{C_1}\varphi \to \mathsf{H}^{D_2}_{C_2}\varphi).$$

Hence, $\vdash \mathsf{H}^{D_1}_{C_1}\varphi \to \mathsf{H}^{D_2}_{C_2}\varphi$ by the Modus Ponens inference rule. □

Lemma 4.2. *If $\varphi_1, \dots, \varphi_n \vdash \psi$, then*

(1) *$\mathsf{K}_C\varphi_1, \dots, \mathsf{K}_C\varphi_n \vdash \mathsf{K}_C\psi$,*
(2) *$\mathsf{H}^{D_1}_{C_1}\varphi_1, \dots, \mathsf{H}^{D_n}_{C_n}\varphi_n \vdash \mathsf{H}^{\bigcup_{i=1}^n D_i}_{\bigcup_{i=1}^n C_i}\psi$, where sets $D_1,\dots,D_n$ are pairwise disjoint.*

Proof. To prove statement (2), apply the deduction lemma for propositional logic $n$ time. Then, $\vdash \varphi_1 \to (\dots \to (\varphi_n \to \psi))$. Thus, $\vdash \mathsf{H}^\varnothing_\varnothing(\varphi_1 \to (\dots \to (\varphi_n \to \psi)))$, by the Strategic Necessitation inference rule. Hence, $\vdash \mathsf{H}^{D_1}_{C_1}\varphi_1 \to \mathsf{H}^{D_1}_{C_1}(\varphi_2 \dots \to (\varphi_n \to \psi))$ by the Cooperation axiom and the Modus Ponens inference rule. Then, $\mathsf{H}^{D_1}_{C_1}\varphi_1 \vdash \mathsf{H}^{D_1}_{C_1}(\varphi_2 \dots \to (\varphi_n \to \psi))$ by the Modus Ponens inference rule. Thus, again by the Cooperation axiom and Modus Ponens, $\mathsf{H}^{D_1}_{C_1}\varphi_1 \vdash \mathsf{H}^{D_2}_{C_2}\varphi_2 \to \mathsf{H}^{D_1\cup D_2}_{C_1\cup C_2}(\varphi_3 \dots \to (\varphi_n \to \psi))$. Therefore, $\mathsf{H}^{D_1}_{C_1}\varphi_1, \dots, \mathsf{H}^{D_n}_{C_n}\varphi_n \vdash \mathsf{H}^{\bigcup_{i=1}^n D_i}_{\bigcup_{i=1}^n C_i}\psi$, by repeating the last two steps $n-2$ times. The proof of the first statement is similar, but it uses the Distributivity axiom instead of the Cooperation axiom. □

## 5 SOUNDNESS

The proof of the soundness of S5 axioms for distributed knowledge (Truth, Negative Introspection, Distributivity, and Monotonicity) is standard. In this section we prove the soundness of each of the remaining axioms as a separate lemma. At the end of this section, Theorem 5.5 states the soundness of the whole system.

Lemma 5.1. *If $w \Vdash \mathsf{H}^{D_1}_{C_1}(\varphi \to \psi)$, $w \Vdash \mathsf{H}^{D_2}_{C_2}\varphi$, and $D_1 \cap D_2 = \varnothing$, then $w \Vdash \mathsf{H}^{D_1\cup D_2}_{C_1\cup C_2}\psi$.*

Proof. Suppose that $w \Vdash \mathsf{H}^{D_1}_{C_1}(\varphi \to \psi)$. Then, by Definition 2.7, there is a strategy profile $\mathbf{s}_1 \in \Delta^{D_1}$ such that for any two epistemic states $w', w''$ and any complete strategy profile $\mathbf{s}'_1$, if $w \sim_{C_1} w'$, $\mathbf{s}_1 =_{D_1} \mathbf{s}'_1$, and $(w', \mathbf{s}'_1, w'') \in M$, then $w'' \Vdash \varphi \to \psi$. Similarly, assumption $w \Vdash \mathsf{H}^{D_2}_{C_2}\varphi$ implies that there is a strategy profile $\mathbf{s}_2 \in \Delta^{D_2}$ such that for any two epistemic states $w', w''$ and any complete strategy profile $\mathbf{s}'_2$, if $w \sim_{C_2} w'$, $\mathbf{s}_2 =_{D_2} \mathbf{s}'_2$, and $(w', \mathbf{s}'_2, w'') \in M$, then $w'' \Vdash \varphi$.

Define a strategy profile $\mathbf{s}$ of coalition $D_1 \cup D_2$ as follows:

$$(\mathbf{s})_a = \begin{cases} (\mathbf{s}_1)_a, & \text{if } a \in D_1, \\ (\mathbf{s}_2)_a, & \text{if } a \in D_2. \end{cases}$$

The strategy profile $\mathbf{s}$ is well-defined because $D_1 \cap D_2 = \varnothing$.

Consider any states $w', w'' \in W$ and any complete strategy profile $\mathbf{s}'$ such that $w \sim_{C_1\cup C_2} w'$, $\mathbf{s} =_{D_1\cup D_2} \mathbf{s}'$ and $(w', \mathbf{s}', w'') \in M$. By Definition 2.7, it suffices to show that $w'' \Vdash \psi$. Indeed, by Definition 2.2, assumption $w \sim_{C_1\cup C_2} w'$ implies that $w \sim_{C_1} w'$ and $w \sim_{C_2} w'$. At the same time, by Definition 2.4, $\mathbf{s} =_{D_1\cup D_2} \mathbf{s}'$ implies that $\mathbf{s} =_{D_1} \mathbf{s}'$ and $\mathbf{s} =_{D_2} \mathbf{s}'$. Thus, $w'' \Vdash \varphi \to \psi$ and $w'' \Vdash \varphi$ by the choice of strategies $\mathbf{s}_1$ and $\mathbf{s}_2$. Therefore, $w'' \Vdash \psi$ by Definition 2.7. □

Lemma 5.2. *If $w \Vdash \mathsf{H}^D_C\varphi$, then $w \Vdash \mathsf{K}_C\mathsf{H}^D_C\varphi$.*

Proof. Suppose that $w \Vdash \mathsf{H}^D_C\varphi$. Then, by Definition 2.7, there is a strategy profile $\mathbf{s} \in \Delta^D$ such that for any two epistemic states $w', w''$ and any complete strategy profile $\mathbf{s}'$, if $w \sim_C w'$, $\mathbf{s} =_D \mathbf{s}'$, and $(w', \mathbf{s}', w'') \in M$, then $w'' \Vdash \varphi$.

Consider any $u \in W$ such that $w \sim_C u$. By Definition 2.7, it suffices to show that $u \Vdash \mathsf{H}^D_C\varphi$. Next, consider any epistemic states $u', u''$ and any complete strategy profile $\mathbf{s}'$ such that $u \sim_C u'$, $\mathbf{s} =_D \mathbf{s}'$, and $(u', \mathbf{s}', u'') \in M$. Again by Definition 2.7, it suffices to show that $u'' \Vdash \varphi$. Indeed, note that $w \sim_C u$ and $u \sim_C u'$ imply that $w \sim_C u'$ by Lemma 2.3. Hence, $u'' \Vdash \varphi$ by the choice of the strategy profile $\mathbf{s}$. □

Lemma 5.3. *If $w \Vdash \mathsf{K}_\varnothing\varphi$, then $w \Vdash \mathsf{H}^\varnothing_\varnothing\varphi$.*

Proof. Suppose $w \Vdash \mathsf{K}_\varnothing\varphi$. Let $\mathbf{s}$ be the unique element of the set $\Delta^\varnothing$. Since $\Delta^\varnothing$ is the set of all functions from $\varnothing$ to $\Delta$, assuming functions are defined as sets of pairs, formally $\mathbf{s}$ is the empty set.

Consider any two epistemic states $w', w''$ and any complete strategy profile $\mathbf{s}'$ such that $w \sim_\varnothing w'$, $\mathbf{s} =_\varnothing \mathbf{s}'$, and $(w', \mathbf{s}', w'') \in M$. By Definition 2.7, it suffices to show that $w'' \Vdash \varphi$. Indeed, note that $w \sim_\varnothing w''$ by Definition 2.2. It then follows from assumption $w \Vdash \mathsf{K}_\varnothing\varphi$ and Definition 2.7 that $w'' \Vdash \varphi$. □

Lemma 5.4. *If $w \Vdash \mathsf{K}_A\mathsf{H}_B^\varnothing\varphi$, then $w \Vdash \mathsf{H}_A^\varnothing\varphi$.*

Proof. Let $\mathbf{s}$ be the unique element of the set $\Delta^\varnothing$. Consider any two epistemic states $w', w''$ and any complete strategy profile $\mathbf{s}'$ such that $w \sim_A w'$, $\mathbf{s} =_\varnothing \mathbf{s}'$, and $(w', \mathbf{s}', w'') \in M$. By Definition 2.7, it suffices to show that $w'' \Vdash \varphi$.

Assumption $w \Vdash \mathsf{K}_A\mathsf{H}_B^\varnothing\varphi$ implies that $w' \Vdash \mathsf{H}_B^\varnothing\varphi$ due to $w \sim_A w'$ and Definition 2.7. Thus, there is a strategy profile $\hat{\mathbf{s}}$ such that for any two epistemic states $u', u''$ and any complete strategy profile $\hat{\mathbf{s}}'$, if $w' \sim_B u'$, $\hat{\mathbf{s}} =_\varnothing \hat{\mathbf{s}}'$, and $(u', \hat{\mathbf{s}}', u'') \in M$, then $u'' \Vdash \varphi$.

Let $u' = w'$, $u'' = w''$, and $\hat{\mathbf{s}}' = \mathbf{s}'$. Note that $w' \sim_B w' = u'$ by Definition 2.2 and that $\hat{\mathbf{s}} =_\varnothing \mathbf{s}'$ by Definition 2.4. Thus, $w'' \Vdash \varphi$. □

Theorem 5.5. *If $\vdash \varphi$, then $w \Vdash \varphi$ for each epistemic state $w$ of each epistemic transition system.* □

## 6 COMPLETENESS

This section contains a proof of the completeness of the logical system. We start with a well-known observation that Positive Introspection principle is provable in S5.

Lemma 6.1. $\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C\mathsf{K}_C\varphi$.

Proof. Formula $\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\varphi$ is an instance of the Negative Introspection axiom. Thus, $\vdash \neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\varphi$ by the law of contrapositive in the propositional logic. Hence, by the Necessitation inference rule, $\vdash \mathsf{K}_C(\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\varphi)$. Thus, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\mathsf{K}_C\varphi. \tag{5}$$

At the same time, $\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \neg\mathsf{K}_C\varphi$ is an instance of the Truth axiom. Thus, $\vdash \mathsf{K}_C\varphi \to \neg\mathsf{K}_C\neg\mathsf{K}_C\varphi$ by contraposition. Hence, taking into account the following instance of the Negative Introspection axiom: $\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi$, one can conclude that $\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi$. The latter, together with statement (5), implies the statement of the lemma by propositional reasoning. □

### 6.1 Harmony

The proof of the completeness is based on the harmony technique proposed by Naumov and Tao [10, 11]. Here, we modify definitions of harmony and complete harmony from [10, 11] to account for the fact that parameters $C$ and $D$ of the modality $\mathsf{H}_C^D$ might be different.

*Definition 6.2.* A pair $(Y, Z)$ is in harmony if $Y \nvdash \mathsf{H}_C^\varnothing\neg\wedge Z'$ for each coalition $C$ and each finite set $Z' \subseteq Z$.

Lemma 6.3. *If pair $(Y, Z)$ is in harmony, then sets $Y$ and $Z$ are consistent.*

Proof. First, suppose that set $Y$ is not consistent. Thus, $Y \vdash \mathsf{H}_\varnothing^\varnothing\neg\wedge\varnothing$. Therefore, by Definition 6.2, pair $(Y, Z)$ is not in harmony.

Next, suppose that set $Z$ is inconsistent. Then, there is a finite set $Z' \subseteq Z$ such that $\vdash \neg\wedge Z'$. Hence, $\vdash \mathsf{H}_\varnothing^\varnothing\neg\wedge Z'$ by the Strategic Necessitation inference rule. Thus, $Y \vdash \mathsf{H}_\varnothing^\varnothing\neg\wedge Z'$. Therefore, by Definition 6.2, pair $(Y, Z)$ is not in harmony. □

Lemma 6.4. *If $X \nvdash \mathsf{H}_C^D\varphi$ and $f$ is an arbitrary function from coalition $D$ to set $\Phi$, then pair $(Y, Z)$ is in harmony, where*

$$\begin{aligned} Y &= \{\psi \mid \mathsf{K}_C\psi \in X\}, \\ Z &= \{\neg\varphi\} \cup \{\chi \mid \mathsf{K}_\varnothing\chi \in X\} \cup \\ &\quad\ \{\tau \mid \mathsf{H}_E^F\tau \in X, F \subseteq D, E \subseteq C, \forall a \in F\,(f(a) = \tau)\}. \end{aligned}$$

Proof. Let pair $(Y, Z)$ not be in harmony. Then, $Y \vdash \mathsf{H}_B^\varnothing\neg\wedge Z'$ for some set $B \subseteq \mathcal{A}$ and some finite set $Z' \subseteq Z$ by Definition 6.2. Thus, by the definition of set $Y$, there are formulae

$$\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \in X \tag{6}$$

such that $\psi_1, \dots, \psi_n \vdash \mathsf{H}_B^\varnothing\neg\wedge Z'$. By Lemma 4.2,

$$\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \vdash \mathsf{K}_C\mathsf{H}_B^\varnothing\neg\wedge Z'.$$

Then, by the Knowledge of Unavoidability axiom

$$\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \vdash \mathsf{H}_C^\varnothing\neg\wedge Z'. \tag{7}$$

Since $Z' \subseteq Z$, by the definition of set $Z$, there are

$$\begin{aligned} &\text{formulae} && \mathsf{K}_\varnothing\chi_1, \dots, \mathsf{K}_\varnothing\chi_m \in X, && (8)\\ &\text{and formulae} && \mathsf{H}_{E_1}^{F_1}\tau_1, \dots, \mathsf{H}_{E_t}^{F_t}\tau_t \in X, && (9)\\ &\text{such that} && F_1, \dots, F_t \subseteq D, E_1, \dots, E_t \subseteq C, && (10)\\ & && \forall i \le t\ \forall a \in F_i\,(f(a) = \tau_i), && (11)\\ &\text{and} && \chi_1, \dots, \chi_m, \tau_1, \dots, \tau_t, \neg\varphi \vdash \wedge Z'. \end{aligned}$$

By the deduction theorem for propositional logic, the last statement implies that $\chi_1, \dots, \chi_m, \tau_1, \dots, \tau_t \vdash \neg\varphi \to \wedge Z'$. Hence, by the law of contrapositive, $\chi_1, \dots, \chi_m, \tau_1, \dots, \tau_t \vdash \neg\wedge Z' \to \varphi$. Then by the Modus Ponens inference rule,

$$\neg\wedge Z', \chi_1, \dots, \chi_m, \tau_1, \dots, \tau_t \vdash \varphi. \tag{12}$$

Without loss of generality, we assume that formulae $\tau_1, \dots, \tau_t$ are pairwise different. Hence, sets $F_1, \dots, F_t$ are pairwise disjoint, by statement (11). Thus, by Lemma 4.2, statement (12) implies that

$$\mathsf{H}_C^\varnothing\neg\wedge Z', \mathsf{H}_\varnothing^\varnothing\chi_1, \dots, \mathsf{H}_\varnothing^\varnothing\chi_m, \mathsf{H}_{E_1}^{F_1}\tau_1, \dots, \mathsf{H}_{E_t}^{F_t}\tau_t \vdash \mathsf{H}_{C\cup E_1\cup\dots\cup E_t}^{F_1\cup\dots\cup F_t}\varphi.$$

Hence, by Lemma 4.1 and due to statement (10),

$$\mathsf{H}_C^\varnothing\neg\wedge Z', \mathsf{H}_\varnothing^\varnothing\chi_1, \dots, \mathsf{H}_\varnothing^\varnothing\chi_m, \mathsf{H}_{E_1}^{F_1}\tau_1, \dots, \mathsf{H}_{E_t}^{F_t}\tau_t \vdash \mathsf{H}_C^D\varphi.$$

Then, by the Empty Coalition axiom,

$$\mathsf{H}_C^\varnothing\neg\wedge Z', \mathsf{K}_\varnothing\chi_1, \dots, \mathsf{K}_\varnothing\chi_m, \mathsf{H}_{E_1}^{F_1}\tau_1, \dots, \mathsf{H}_{E_t}^{F_t}\tau_t \vdash \mathsf{H}_C^D\varphi.$$

Therefore, by statement (7),

$$\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n, \mathsf{K}_\varnothing\chi_1, \dots, \mathsf{K}_\varnothing\chi_m, \mathsf{H}_{E_1}^{F_1}\tau_1, \dots, \mathsf{H}_{E_t}^{F_t}\tau_t \vdash \mathsf{H}_C^D\varphi.$$

It then follows from statements (6), (8), and (9) that $X \vdash \mathsf{H}_C^D\varphi$. □

Lemma 6.5. *For any pair $(Y, Z)$ in harmony, any formula $\varphi \in \Phi$, and any set $C \subseteq \mathcal{A}$, either pair $(Y \cup \{\neg\mathsf{H}_C^\varnothing\varphi\}, Z)$ or pair $(Y, Z \cup \{\varphi\})$ is in harmony.*

Proof. Suppose that neither pair $(Y \cup \{\neg \mathsf{H}^{\varnothing}_C \varphi\}, Z)$ nor pair $(Y, Z \cup \{\varphi\})$ is in harmony. Then, by Definition 6.2, there are sets $E_1, E_2 \subseteq \mathcal{A}$ and finite sets $Z_1 \subseteq Z$ and $Z_2 \subseteq Z \cup \{\varphi\}$ such that $Y, \neg \mathsf{H}^{\varnothing}_C \varphi \vdash \mathsf{H}^{\varnothing}_{E_1} \neg \wedge Z_1$ and $Y \vdash \mathsf{H}^{\varnothing}_{E_2} \neg \wedge Z_2$. Let $D = C \cup E_1 \cup E_2$. Then, by Lemma 4.1 applied three times, we have

$$Y, \neg \mathsf{H}^{\varnothing}_D \varphi \vdash \mathsf{H}^{\varnothing}_D \neg \wedge Z_1 \tag{13}$$

and

$$Y \vdash \mathsf{H}^{\varnothing}_D \neg \wedge Z_2. \tag{14}$$

Let $Z_0 = Z_1 \cup (Z_2 \setminus \{\varphi\})$. Then, formulae $\wedge Z_0 \to \wedge Z_1$ and $\varphi \wedge (\wedge Z_0) \to \wedge Z_2$ are propositional tautologies because $Z_1 \subseteq Z_0$ and $Z_2 \subseteq Z_0 \cup \{\varphi\}$. Thus, by the law of contrapositive, $\vdash \neg \wedge Z_1 \to \neg \wedge Z_0$ and $\vdash \neg \wedge Z_2 \to \neg(\varphi \wedge (\wedge Z_0))$. Hence, $\vdash \neg \wedge Z_1 \to \neg \wedge Z_0$ and $\vdash \neg \wedge Z_2 \to (\varphi \to \neg \wedge Z_0)$. Therefore, $\neg \wedge Z_1 \vdash \neg \wedge Z_0$ and $\neg \wedge Z_2 \vdash \varphi \to \neg \wedge Z_0$ by the Modus Ponens inference rule. Thus, by Lemma 4.2,

$$\mathsf{H}^{\varnothing}_D \neg \wedge Z_1 \vdash \mathsf{H}^{\varnothing}_D \neg \wedge Z_0,$$
$$\mathsf{H}^{\varnothing}_D \neg \wedge Z_2 \vdash \mathsf{H}^{\varnothing}_D (\varphi \to \neg \wedge Z_0).$$

Then, due to statements (13) and (14),

$$Y, \neg \mathsf{H}^{\varnothing}_D \varphi \vdash \mathsf{H}^{\varnothing}_D \neg \wedge Z_0, \tag{15}$$

$$Y \vdash \mathsf{H}^{\varnothing}_D (\varphi \to \neg \wedge Z_0). \tag{16}$$

By the Cooperation axiom and the Modus Ponens inference rule, formula (16) implies that $Y, \mathsf{H}^{\varnothing}_D \varphi \vdash \mathsf{H}^{\varnothing}_D \neg \wedge Z_0$. Hence, $Y \vdash \mathsf{H}^{\varnothing}_D \neg \wedge Z_0$ by the laws of propositional reasoning from statement (15). The last statement, by Definition 6.2, contradicts the assumption that pair $(Y, Z)$ is in harmony because $Z_0 \subseteq Z$. □

## 6.2 Complete Harmony

*Definition 6.6.* A pair in harmony $(Y, Z)$ is in *complete* harmony if for each $\varphi \in \Phi$ and each coalition $C$, either $\neg \mathsf{H}^{\varnothing}_C \varphi \in Y$ or $\varphi \in Z$.

Lemma 6.7. *If pair $(Y, Z)$ is in harmony, then there is a pair in complete harmony $(Y', Z')$, where $Y \subseteq Y'$ and $Z \subseteq Z'$.*

Proof. Recall that the set of agent $\mathcal{A}$ and the set of propositional variables are countable. Thus, the set of all formulae $\Phi$ is also countable. Let sequence $\mathsf{H}^{\varnothing}_{C_1} \varphi_1, \mathsf{H}^{\varnothing}_{C_2} \varphi_2, \dots$ be an enumeration of the set $\{\mathsf{H}^{\varnothing}_C \varphi \mid \varphi \in \Phi, C \subseteq \mathcal{A}\}$. We define two chains of sets $Y_1 \subseteq Y_2 \subseteq \dots$ and $Z_1 \subseteq Z_2 \subseteq \dots$ such that pair $(Y_n, Z_n)$ is in harmony for each $n \geq 1$. These two chains are defined recursively as follows:

(1) $Y_1 = Y$ and $Z_1 = Z$,
(2) if pair $(Y_n, Z_n)$ is in harmony, then, by Lemma 6.5, either pair $(Y_n \cup \{\neg \mathsf{H}^{\varnothing}_{C_n} \varphi_n\}, Z_n)$ or pair $(Y_n, Z_n \cup \{\varphi_n\})$ is in harmony. Let $(Y_{n+1}, Z_{n+1})$ be $(Y_n \cup \{\neg \mathsf{H}^{\varnothing}_{C_n} \varphi_n\}, Z_n)$ in the former case and $(Y_n, Z_n \cup \{\varphi_n\})$ in the latter case.

Let $Y' = \bigcup_n Y_n$ and $Z' = \bigcup_n Z_n$. Note that $Y = Y_1 \subseteq Y'$ and $Z = Z_1 \subseteq Z'$.

We next show that pair $(Y', Z')$ is in harmony. Suppose the opposite. Then, by Definition 6.2, there is a coalition $C$ and a finite set $Z'' \subseteq Z'$ such that $Y' \vdash \mathsf{H}^{\varnothing}_C \neg \wedge Z''$. Since a deduction uses only finitely many assumptions, there exists an integer $n_1 \geq 1$ such that

$$Y_{n_1} \vdash \mathsf{H}^{\varnothing}_C \neg \wedge Z''. \tag{17}$$

At the same time, since set $Z''$ is finite, there must exist an integer $n_2 \geq 1$ such that $Z'' \subseteq Z_{n_2}$. Let $n = \max\{n_1, n_2\}$. Note that $\neg \wedge Z'' \vdash \neg \wedge Z_n$ because $Z'' \subseteq Z_{n_2} \subseteq Z_n$. Thus, $\mathsf{H}^{\varnothing}_C \neg \wedge Z'' \vdash \mathsf{H}^{\varnothing}_C \neg \wedge Z_n$ by Lemma 4.2. Hence, $Y_{n_1} \vdash \mathsf{H}^{\varnothing}_C \neg \wedge Z_n$ due to statement (17). Thus, $Y_n \vdash \mathsf{H}^{\varnothing}_C \neg \wedge Z_n$ because $Y_{n_1} \subseteq Y_n$. Then, pair $(Y_n, Z_n)$ is not in harmony, which contradicts the choice of pair $(Y_n, Z_n)$. Therefore, pair $(Y', Z')$ is in harmony.

We finally show that pair $(Y', Z')$ is in complete harmony. Indeed, consider any formula $\mathsf{H}^{\varnothing}_C \varphi \in \Phi$. Since sequence $\mathsf{H}^{\varnothing}_{C_1} \varphi_1, \mathsf{H}^{\varnothing}_{C_2} \varphi_2, \dots$ is an enumeration of all formulae in the set $\{\mathsf{H}^{\varnothing}_C \varphi \mid \varphi \in \Phi, C \subseteq \mathcal{A}\}$, there must exist an integer $k \geq 1$ such that $\mathsf{H}^{\varnothing}_C \varphi = \mathsf{H}^{\varnothing}_{C_k} \varphi_k$. Then, by the choice of pair $(Y_{k+1}, Z_{k+1})$, either $\neg \mathsf{H}^{\varnothing}_C \varphi = \neg \mathsf{H}^{\varnothing}_{C_k} \varphi_k \in Y_{k+1} \subseteq Y'$ or $\varphi = \varphi_k \in Z_{k+1} \subseteq Z'$. Therefore, the pair $(Y', Z')$ is in complete harmony. □

## 6.3 Canonical Epistemic Transition System

In this section, we use the "unravelling" technique [15] to define a canonical transition system $ETS(X_0) = (W, \{\sim_a\}_{a \in \mathcal{A}}, \Delta, M, \pi)$ for an arbitrary maximal consistent set of formulae $X_0 \subseteq \Phi$.

*Definition 6.8.* The set of epistemic states $W$ consists of all finite sequences $X_0, C_1, X_1, C_2, \dots, C_n, X_n$, such that

(1) $n \geq 0$,
(2) $X_i$ is a maximal consistent subset of $\Phi$ for each $i \geq 1$,
(3) $C_i$ is a coalition for each $i \geq 1$,
(4) $\{\varphi \mid \mathsf{K}_{C_i} \varphi \in X_{i-1}\} \subseteq X_i$ for each $i \geq 1$.

Set $W$ can be viewed as a tree whose nodes are labeled with maximal consistent sets and whose edges are labeled with coalitions. For any sequence $x = x_1, \dots, x_n$ and an element $y$, by sequence $x :: y$ we mean $x_1, \dots, x_n, y$. If sequence $x$ is nonempty, then by $hd(x)$ we mean element $x_n$. The abbreviation $hd$ stands for "head".

*Definition 6.9.* For any state $w = X_0, C_1, \dots, C_n, X_n$ and any state $w' = X_0, C'_1, \dots, C'_m, X'_m$, let $w \sim_a w'$ if there is $k$ such that

(1) $0 \leq k \leq \min\{n, m\}$,
(2) $X_i = X'_i$ for each $i$ such that $1 \leq i \leq k$,
(3) $C_i = C'_i$ for each $i$ such that $1 \leq i \leq k$,
(4) $a \in C_i$ for each $i$ such that $k < i \leq n$,
(5) $a \in C'_i$ for each $i$ such that $k < i \leq m$.

*Definition 6.10.* $\Delta = \{(\varphi, C, [w]_C) \mid \varphi \in \Phi, C \subseteq \mathcal{A}, w \in W\}$.

Informally, each action $(\varphi, C, [w]_C) \in \Delta$ of an agent $c \in C$ consists of a formula $\varphi$ that coalition $C$ is trying to achieve and an indistinguishability class $[w]_C$. Class $[w]_C$ acts as a "signature" with which coalition $C$ "signs" its action. As per definition of the mechanism $M$ below, an action might have an effect only if it is signed with the right signature.

*Definition 6.11.* For any states $w, w' \in W$ and any complete strategy profile $\mathbf{s} \in \Phi^{\mathcal{A}}$, let $(w, \mathbf{s}, w') \in M$ if set $hd(w')$ contains all elements of the set

$$\{\varphi \mid (\mathsf{H}^D_C \varphi \in hd(w)) \wedge \forall a \in D((\mathbf{s})_a = (\varphi, C, [w]_C))\}.$$

*Definition 6.12.* $\pi(p) = \{w \in W \mid p \in hd(w)\}$.

This concludes the definition of the canonical epistemic transition system $ETS(X_0)$.

## 6.4 Properties of the Canonical System

Lemma 6.13. *For any state $X_0, C_1, X_1, \dots, C_n, X_n \in W$ and any integer $k \le n$, if $\mathsf{K}_C\varphi \in X_n$ and $C \subseteq C_i$ for each integer $i$ such that $k < i \le n$, then $\mathsf{K}_C\varphi \in X_k$.*

Proof. Suppose that there is $k \le n$ such that $\mathsf{K}_C\varphi \notin X_k$. Let $m$ be the maximal such $k$. Note that $m < n$ due to the assumption $\mathsf{K}_C\varphi \in X_n$ of the lemma. Thus, $m + 1 \le n$.

Assumption $\mathsf{K}_C\varphi \notin X_m$ implies $\neg\mathsf{K}_C\varphi \in X_m$ due to the maximality of the set $X_m$. Hence, $X_m \vdash \mathsf{K}_C\neg\mathsf{K}_C\varphi$ by the Negative Introspection axiom. Thus, $X_m \vdash \mathsf{K}_{C_{m+1}}\neg\mathsf{K}_C\varphi$ by the Monotonicity axiom and the assumption $C \subseteq C_{m+1}$ of the lemma (recall that $m + 1 \le n$). Then, $\mathsf{K}_{C_{m+1}}\neg\mathsf{K}_C\varphi \in X_m$ due to the maximality of the set $X_m$. Hence, $\neg\mathsf{K}_C\varphi \in X_{m+1}$ by Definition 6.8. Thus, $\mathsf{K}_C\varphi \notin X_{m+1}$ due to the consistency of the set $X_{m+1}$, which is a contradiction with the choice of integer $m$. ◻

Lemma 6.14. *For any state $X_0, C_1, X_1, \dots, C_n, X_n \in W$ and any integer $k \le n$, if $\mathsf{K}_C\varphi \in X_k$ and $C \subseteq C_i$ for each integer $i$ such that $k < i \le n$, then $\varphi \in X_n$.*

Proof. We prove the lemma by induction on the distance between $n$ and $k$. In the base case $n = k$, the assumption $\mathsf{K}_C\varphi \in X_n$ implies $X_n \vdash \varphi$ by the Truth axiom. Therefore, $\varphi \in X_n$ due to the maximality of set $X_n$.

Suppose that $k < n$. Assumption $\mathsf{K}_C\varphi \in X_k$ implies $X_k \vdash \mathsf{K}_C\mathsf{K}_C\varphi$ by Lemma 6.1. Thus, $X_k \vdash \mathsf{K}_{C_{k+1}}\mathsf{K}_C\varphi$ by the Monotonicity axiom, the condition $k < n$ of the inductive step, and the assumption $C \subseteq C_{k+1}$ of the lemma. Then, $\mathsf{K}_{C_{k+1}}\mathsf{K}_C\varphi \in X_k$ by the maximality of set $X_k$. Hence, $\mathsf{K}_C\varphi \in X_{k+1}$ by Definition 6.8. Therefore, $\varphi \in X_n$ by the induction hypothesis. ◻

The next lemma follows from Lemma 6.13, Lemma 6.14, and Definition 6.9 because there is a unique path between any two nodes in a tree.

Lemma 6.15. *For any epistemic states $w, w' \in W$ such that $w \sim_C w'$, if $\mathsf{K}_C\varphi \in hd(w)$, then $\varphi \in hd(w')$.* ◻

For any triple $v = (x, y, z)$, by $pr_1(v)$, $pr_2(v)$, and $pr_3(v)$ we mean $x$, $y$, and $z$, respectively.

Lemma 6.16. *If $w_1, w_2 \in pr_3(v)$, then $w_1 \sim_{pr_2(v)} w_2$, for each $w_1, w_2 \in W$ and each $v \in \Delta$.*

Proof. Let $v = (\varphi, C, [w]_C)$, where $\varphi \in \Phi$, $C \subseteq \mathcal{A}$, and $w \in W$. Assumption $w_1, w_2 \in pr_3(v)$ implies that $w_1, w_2 \in [w]_C$. Thus, $w_1 \sim_C w \sim_C w_2$. Hence, $w_1 \sim_C w_2$. Therefore, $w_1 \sim_{pr_2(v)} w_2$. ◻

Lemma 6.17. *For any state $w \in W$ if $\neg\mathsf{K}_C\varphi \in hd(w)$, then there is a state $w' \in W$ such that $w \sim_C w'$ and $\neg\varphi \in hd(w')$.*

Proof. Consider the set of formulae

$$X = \{\neg\varphi\} \cup \{\psi \mid \mathsf{K}_C\psi \in hd(w)\}.$$

First, we show that set $X$ is consistent. Assume the opposite. Then, there are $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \in hd(w)$ such that $\psi_1, \dots, \psi_n \vdash \varphi$. Thus, $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \vdash \mathsf{K}_C\varphi$ by Lemma 4.2. Therefore, $hd(w) \vdash \mathsf{K}_C\varphi$ by the choice of formulae $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n$, which contradicts the consistency of set $hd(w)$ due to the assumption $\neg\mathsf{K}_C\varphi \in hd(w)$.

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 6.8 and the choice of set $X$. Furthermore, $w \sim_C w'$ by Definition 6.9. To finish the proof, note that $\neg\varphi \in X \subseteq \hat{X} = hd(w')$ by the choice of $X$. ◻

Lemma 6.18. *Let $w, w', u \in W$ be epistemic states, $\mathsf{H}^D_C\varphi \in hd(w)$ be a formula, and $\mathbf{s}$ be a complete strategy profile such that $(\mathbf{s})_a = (\varphi, C, [w]_C)$ for each $a \in D$. If $w \sim_C w'$ and $(w', \mathbf{s}, u) \in M$, then $\varphi \in hd(u)$.*

Proof. Let $\mathsf{H}^D_C\varphi \in hd(w)$. Then, $hd(w) \vdash \mathsf{K}_C\mathsf{H}^D_C\varphi$ by the Strategic Introspection axiom. Thus, $\mathsf{K}_C\mathsf{H}^D_C\varphi \in hd(w)$ by the maximality of the set $hd(w)$. Hence, $\mathsf{H}^D_C\varphi \in hd(w')$ by Lemma 6.15 and the assumption $w \sim_C w'$. Then, $\varphi \in hd(u)$ by Definition 6.11 because of assumption $(w', \mathbf{s}, u) \in M$ and assumption $(\mathbf{s})_a = (\varphi, C, [w]_C)$ for each $a \in D$. ◻

Lemma 6.19. *For any state $w \in W$, any formula $\neg\mathsf{H}^D_C\varphi \in hd(w)$, and any strategy profile $\mathbf{s} \in \Delta^D$, there are epistemic states $w', u \in W$ and a complete strategy profile $\mathbf{s}'$ such that $w \sim_C w'$, $\mathbf{s} =_D \mathbf{s}'$, $(w', \mathbf{s}', u) \in M$, and $\varphi \notin hd(u)$.*

Proof. Set $D_0 = \{a \in D \mid pr_2((\mathbf{s})_a) \nsubseteq C\}$ is finite because set $D$ is finite by Definition 2.6. Let $D_0 = \{a_1, \dots, a_n\}$.

Next, consider sets

$$\begin{aligned} Y &= \{\psi \mid \mathsf{K}_C\psi \in hd(w)\} \\ Z &= \{\neg\varphi\} \cup \{\chi \mid \mathsf{K}_\varnothing\chi \in hd(w)\} \cup \\ &\quad \{\tau \mid \mathsf{H}^F_E\tau \in hd(w), F \subseteq D, E \subseteq C, \forall a \in F\,(pr_1((\mathbf{s})_a) = \tau)\}. \end{aligned}$$

By Lemma 6.4, pair $(Y, Z)$ is in harmony. Thus, by Lemma 6.7, there is a pair $(Y', Z')$ in complete harmony such that $Y \subseteq Y'$ and $Z \subseteq Z'$. By Lemma 6.3, sets $Y'$ and $Z'$ are consistent. Let $Y''$ and $Z''$ be any maximal consistent extensions of sets $Y'$ and $Z'$ respectively.

Recall that $n$ is the number of elements in set $D_0$. Define sequences $w_1, \dots, w_{n+1}$ as follows:

$$\begin{aligned} w_1 &= w :: C :: Y'' \\ w_2 &= w :: C :: Y'' :: C :: Y'' \\ w_3 &= w :: C :: Y'' :: C :: Y'' :: C :: Y'' \\ \dots & \quad \dots \\ w_{n+1} &= w \underbrace{:: C :: Y'' :: C :: Y'' \cdots :: C :: Y''}_{::C::Y'' \text{ repeated } n+1 \text{ times}}. \end{aligned}$$

Claim 1. *$w_k \in W$ where $1 \le k \le n + 1$.*

Proof of Claim. We prove the claim by induction on integer $k$. If $k = 1$, then, by Definition 6.8, it suffices to show that $\{\varphi \mid \mathsf{K}_C\varphi \in hd(w)\} \subseteq Y''$. Indeed, by the choice of $Y, Y', Y''$, we have $\{\varphi \mid \mathsf{K}_C\varphi \in hd(w)\} = Y \subseteq Y' \subseteq Y''$.

For the induction step, by Definition 6.8 and the definition of $w_1, \dots, w_{n+1}$, it suffices to show that $\{\varphi \mid \mathsf{K}_C\varphi \in Y''\} \subseteq Y''$. This follows from the Truth axiom and the maximality of set $Y''$. ◻

Claim 2. *$w \sim_C w_i$ for each $i \le n + 1$.*

Proof of Claim. The claims follows from Definition 2.2, Definition 6.9, and the definition of $w_1, \dots, w_{n+1}$. ◻

Claim 3. *If $w_i \sim_E w_j$ and $i \neq j$, then $E \subseteq C$.*

Proof of Claim. Consider any agent $a \in E$. Assumption $w_i \sim_E w_j$ implies that $w_i \sim_a w_j$ by Definition 2.2. Thus, $a \in C$ by Definition 6.9, the definition of $w_i$ and $w_j$, and the assumption $i \neq j$. $\square$

Claim 4. *There is an integer $k \leq n+1$ such that for each $i \leq n$, we have $w_k \notin pr_3((s)_{a_i})$.*

Proof of Claim. Suppose that for each $k \leq n+1$ there is $i \leq n$ such that $w_k \in pr_3((s)_{a_i})$. Then, by the pigeonhole principle, there is $j \leq n$ and $k_1, k_2 \leq n+1$ such that $k_1 \neq k_2$ and $w_{k_1}, w_{k_2} \in pr_3((s)_{a_j})$. Hence, $w_{k_1} \sim_{pr_2((s)_{a_j})} w_{k_2}$ by Lemma 6.16. Hence, $pr_2((s)_{a_j}) \subseteq C$ by Claim 3, which contradicts the choice of set $D_0$. $\square$

We now continue with the proof of the lemma. Choose $w' \in \{w_1, \dots, w_{n+1}\}$ such that

$$\forall a \in D_0\ (w' \notin pr_3((s)_a)). \tag{18}$$

Such $w_k$ exists by Claim 4 and the choice of agents $a_1, \dots, a_n$. Let $u$ be sequence $w :: \varnothing :: Z''$. Note that $u \in W$ by Definition 6.8.

Let $\mathbf{s}'$ be the complete strategy profile such that

$$(\mathbf{s}')_a = \begin{cases} (\mathbf{s})_a, & a \in D, \\ (\top, C, [w]_C), & \text{otherwise.} \end{cases} \tag{19}$$

Note that $\mathbf{s} =_D \mathbf{s}'$ by Definition 2.4.

Claim 5. $(w', \mathbf{s}', u) \in M$.

Proof of Claim. Consider any formula $\mathsf{H}_E^F \tau \in hd(w')$ such that

$$(\mathbf{s}')_a = (\tau, E, [w']_E), \text{ for each } a \in F. \tag{20}$$

By Definition 6.11, it suffices to show that $\tau \in hd(u)$. We consider the following four cases separately.

*Case 1:* $F = \varnothing$. Thus, either $\neg\mathsf{H}_E^F \tau \in Y' \subseteq Y'' = hd(w')$ or $\tau \in Z' \subseteq Z'' = hd(u)$ by Definition 6.6, because pair $(Y', Z')$ is in complete harmony. Hence, $\tau \in hd(u)$ because set $hd(w')$ is consistent and $\mathsf{H}_E^F \tau \in hd(w')$.

*Case 2:* $F \nsubseteq D$. Then, there is an agent $f_0 \in F$ such that $f_0 \notin D$. Hence, $(\mathbf{s}')_{f_0} = (\tau, E, [w']_E)$ by equation (20). At the same time, $(\mathbf{s}')_{f_0} = (\top, C, [w]_C)$ by equation (19). Thus, $\tau = \top$. Therefore, $\tau \in hd(u)$ because $hd(u)$ is a maximal consistent set.

*Case 3:* $F \neq \varnothing$ and $F \subseteq D_0$. Consider any $f_0 \in F$. Note that $w' \notin pr_3((s)_{f_0})$ by equation (18) because $f_0 \in F \subseteq D_0$. Therefore, $w' \notin [w']_E$ by equation (20), which contradicts Lemma 2.3.

*Case 4:* $F \subseteq D$ and $F \nsubseteq D_0$. Then, there is an agent $f_0 \in F$ such that $f_0 \in D \setminus D_0$. Hence, $pr_2((\mathbf{s})_{f_0}) \subseteq C$ by the definition of set $D_0$. Thus, $E \subseteq C$ due to equation (20).

Recall that $\mathsf{H}_E^F \tau \in hd(w')$ by the choice of formula $\mathsf{H}_E^F \tau$. Hence, $hd(w') \vdash \mathsf{K}_E \mathsf{H}_E^F \tau$ by the Strategic Introspection axiom. Hence, $hd(w') \vdash \mathsf{K}_C \mathsf{H}_E^F \tau$ by the Monotonicity axiom because $E \subseteq C$. Thus, $\mathsf{K}_C \mathsf{H}_E^F \tau \in hd(w')$ because set $hd(w')$ is a maximal consistent set. Then, $\mathsf{H}_E^F \tau \in hd(w)$ by Lemma 6.15 and Claim 2. Therefore, $\tau \in Z \subseteq Z' \subseteq Z'' = hd(u)$ by the choice of set $Z$, equation (20), and because $F \subseteq D$ and $E \subseteq C$. $\square$

To finish the proof of the lemma, notice that $\neg\varphi \in Z \subseteq Z' \subseteq Z'' = hd(u)$ by the choice of set $Z$. Therefore, $\varphi \notin hd(u)$ because set $hd(u)$ is consistent. $\square$

Lemma 6.20. *$w \Vdash \varphi$ iff $\varphi \in hd(w)$ for each epistemic state $w \in W$ and each formula $\varphi \in \Phi$.*

Proof. We prove the lemma by induction on the structural complexity of formula $\varphi$. If formula $\varphi$ is a propositional variable, then the required follows from Definition 2.7 and Definition 6.12. The cases of formula $\varphi$ being a negation or an implication follow from Definition 2.7, and the maximality and the consistency of the set $hd(w)$ in the standard way.

Let formula $\varphi$ have the form $\mathsf{K}_C \psi$.

($\Rightarrow$) Suppose that $\mathsf{K}_C \psi \notin hd(w)$. Then, $\neg\mathsf{K}_C \psi \in hd(w)$ by the maximality of set $hd(w)$. Thus, by Lemma 6.17, there is $w' \in W$ such that $w \sim_C w'$ and $\neg\psi \in hd(w')$. By the consistency of $hd(w')$, we have $\psi \notin hd(w')$. Hence, $w' \nVdash \psi$ by the induction hypothesis. Therefore, $w \nVdash \mathsf{K}_C \psi$ by Definition 2.7.

($\Leftarrow$) Assume that $\mathsf{K}_C \psi \in hd(w)$. Consider any $w' \in W$ such that $w \sim_C w'$. By Definition 2.7, it suffices to show that $w' \Vdash \psi$. Indeed, $\psi \in hd(w')$ by Lemma 6.15. Therefore, by the induction hypothesis, $w' \Vdash \psi$.

Let formula $\varphi$ have the form $\mathsf{H}_C^D \psi$.

($\Rightarrow$) Suppose that $\mathsf{H}_C^D \psi \notin hd(w)$. Then, $\neg\mathsf{H}_C^D \psi \in hd(w)$ due to the maximality of the set $hd(w)$. Hence, by Lemma 6.19, for any strategy profile $\mathbf{s} \in \Delta^D$, there are epistemic states $w', w'' \in W$ and a complete strategy profile $\mathbf{s}'$ such that $w \sim_C w'$, $\mathbf{s} =_D \mathbf{s}'$, $(w', \mathbf{s}', w'') \in M$, and $\psi \notin hd(w'')$. Thus, $w'' \nVdash \psi$ by the induction hypothesis. Therefore, $w \nVdash \mathsf{H}_C^D \psi$ by Definition 2.7.

($\Leftarrow$) Assume that $\mathsf{H}_C^D \psi \in hd(w)$. Let $\mathbf{s}$ be a strategy profile of coalition $D$ such that $(\mathbf{s})_a = (\psi, C, [w]_C)$ for each agent $a \in D$. Consider any epistemic states $w', w'' \in W$ and a complete strategy profile $\mathbf{s}'$ such that $w \sim_C w'$, $\mathbf{s}' =_D \mathbf{s}$, and $(w', \mathbf{s}, w'') \in M$. By Definition 2.7, it suffices to show that $w'' \Vdash \psi$.

Indeed, $(\mathbf{s}')_a = (\mathbf{s})_a = (\psi, C, [w]_C)$ for each agent $a \in D$ by the choice of $\mathbf{s}$ and because $\mathbf{s}' =_D \mathbf{s}$. Thus, $\varphi \in hd(w'')$ by Lemma 6.18 and due to the assumptions $w \sim_C w'$ and $(w', \mathbf{s}, w'') \in M$. Therefore, $w'' \Vdash \psi$ by the induction hypothesis. $\square$

## 6.5 Completeness: the Final Step

Theorem 6.21. *If $w \Vdash \varphi$ for each epistemic state $w$ of each epistemic transition system, then $\vdash \varphi$.*

Proof. Suppose that $\nvdash \varphi$. Then let $X_0$ be a maximal consistent set such that $\neg\varphi \in X_0$. Consider the canonical transition system $ETS(X_0)$ defined in Section 6.3. Let $w$ be the single-element sequence $X_0$. Then, $w \in W$ by Definition 6.8. Thus, $w \Vdash \neg\varphi$ by Lemma 6.20. Therefore, $w \nVdash \varphi$ by Definition 2.7. $\square$

# 7 CONCLUSION

In this paper, building on the existing body of literature on know-how strategies, we introduced a notion of a second-order know-how strategy and presented a sound and complete axiomatization of the interplay between the distributed knowledge modality and the second-order know-how modality. The logical system includes a new principle, the Knowledge of Unavoidability, that was not present in any of the existing axiomatizations of different forms of the first-order know-how modality. The completeness proof is based on a harmony technique [10] which was not necessary to prove the completeness of bimodal logics of first-order know-how.

## REFERENCES


[1] Thomas Ågotnes and Natasha Alechina. 2016. Coalition Logic with Individual, Distributed and Common Knowledge. *Journal of Logic and Computation* (2016). https://doi.org/10.1093/logcom/exv085 exv085.

[2] Jan Broersen. 2008. *A Logical Analysis of the Interaction between 'Obligation-to-do' and 'Knowingly Doing'*. Springer Berlin Heidelberg, Berlin, Heidelberg, 140–154.

[3] Jan Broersen, Andreas Herzig, and Nicolas Troquard. 2009. What groups do, can do, and know they can do: an analysis in normal modal logics. *Journal of Applied Non-Classical Logics* 19, 3 (2009), 261–289. https://doi.org/10.3166/jancl.19.261-289

[4] Raul Fervari, Andreas Herzig, Yanjun Li, and Yanjing Wang. 2017. Strategically knowing how. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*. 1031–1038.

[5] Wojciech Jamroga and Thomas Ågotnes. 2007. Constructive knowledge: what agents can achieve under imperfect information. *Journal of Applied Non-Classical Logics* 17, 4 (2007), 423–475. https://doi.org/10.3166/jancl.17.423-475

[6] Wojciech Jamroga and Wiebe van der Hoek. 2004. Agents that know how to play. *Fundamenta Informaticae* 63, 2-3 (2004), 185–219.

[7] C. Kelp and N. J.L.L. Pedersen. 2011. Second-order knowledge. In *The Routledge Companion to Epistemology*, S. Bernecker and D. Pritchard (Eds.). Routledge, London, 586–596.

[8] Pavel Naumov and Jia Tao. 2017. Coalition Power in Epistemic Transition Systems. In *Proceedings of the 2017 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 723–731.

[9] Pavel Naumov and Jia Tao. 2017. Strategic Coalitions with Perfect Recall. *arXiv:1707.04298* (2017).

[10] Pavel Naumov and Jia Tao. 2017. Together We Know How to Achieve: An Epistemic Logic of Know-How. In *16th conference on Theoretical Aspects of Rationality and Knowledge (TARK), July 24-26, 2017, EPTCS 251*. 441–453.

[11] Pavel Naumov and Jia Tao. 2017. Together We Know How to Achieve: An Epistemic Logic of Know-How. *arXiv:1705.09349* (2017).

[12] Pavel Naumov and Jia Tao. 2018. Strategic Coalitions with Perfect Recall. In *Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence*. (to appear).

[13] Marc Pauly. 2001. *Logic for Social Software*. Ph.D. Dissertation. Institute for Logic, Language, and Computation.

[14] Marc Pauly. 2002. A Modal Logic for Coalitional Power in Games. *Journal of Logic and Computation* 12, 1 (2002), 149–166. https://doi.org/10.1093/logcom/12.1.149

[15] Henrik Sahlqvist. 1975. Completeness and correspondence in the first and second order semantics for modal logic. *Studies in Logic and the Foundations of Mathematics* 82 (1975), 110–143. (Proc. of the 3rd Scandinavial Logic Symposium, Uppsala, 1973).

[16] Johan van Benthem. 2001. Games in Dynamic-Epistemic Logic. *Bulletin of Economic Research* 53, 4 (2001), 219–248. https://doi.org/10.1111/1467-8586.00133

[17] Yanjing Wang. 2015. A logic of knowing how. In *Logic, Rationality, and Interaction*. Springer, 392–405. https://doi.org/10.1007/978-3-662-48561-3_32

[18] Yanjing Wang. 2016. A Logic of Goal-directed Knowing How. *Synthese* (2016). (to appear).