Epistemic Logic of Know-Who

Sophia Epstein,[1] **Pavel Naumov**[2]
[1] Claremont McKenna College
[2] King's College
sepstein22@cmc.edu, pgn2@cornell.edu

Abstract

The paper suggests a definition of "know who" as a modality using Grove-Halpern semantics of names. It also introduces a logical system that describes the interplay between modalities "knows who", "knows", and "for all agents". The main technical result is a completeness theorem for the proposed system.

Introduction

The ability of artificial agents to properly identify humans and other machines is critical in many AI applications from online and checkout-less shopping, robotic nurses, and unmanned aircraft systems to security, law-enforcement, and lethal autonomous weaponry. Most of the current systems rely on physical identifiers such as facial images, fingerprints, signatures, government-issued IDs, iris recognition, credit card security chips, passwords, and radio signals. Knowing one of these identifiers does not imply knowing the others or knowing who the person (or a machine) "really" is. In this paper we propose a formal framework for defining and reasoning about "knowing (somebody) who".

The Night Stalker

On July 27th, 1981, a real estate agent came to see a house for sale near Santa Barbara, California. Inside the house at 449 Toltec Way in Goleta, the agent found the bodies of 35-year-old Cheri Domingo, who was house-sitting the place, and of her former boyfriend, 27-year-old Gregory Sanchez (Hardy 1981). Within several days, the Santa Barbara County sheriff's spokesman Russ Birchim announced that the police knew who the killer was. He was the same man who committed a nonfatal knife attack on another couple in the same neighborhood 22 months ago. Birchim said that the deputies dubbed the killer "Night Stalker" (Hurst 1981).

Did Birchim really *know* who the murderer was? It took almost 40 years for the police to find out that "Night Stalker" is actually "East Area Rapist" who raped 50 people in Northern California in the 1970s, almost 400 miles away from Santa Barbara. The same person was also known as "Visalia Ransacker" and "Golden State Killer". It also was discovered that the same person was known to California police as sergeant DeAngelo serving in Auburn, California police forces from August 1976 to July 1979, when he was arrested and sentenced to six months probation for shoplifting a hammer and dog repellent (Levine 2018; Serna and Oreskes 2018). Did the sentencing judge know who the shoplifter *really* was?

As the example above shows, the same person might be known under different names and knowing one of the person's names does not necessarily imply knowing all of them. To define the meaning of "know who" one needs to fix a name space. Knowing the person under one name space does not imply knowing the same person under another. For example, Birchim knew who the murderer was using a hypothetical name space consisting of "Night Stalker", "Morning Stalker", "Day Stalker", and "Evening Stalker", but did not know the murderer in a hypothetical space "East Area Rapist", "North Area Rapist", "West Area Rapist", and "South Area Rapist".

There are many other real-world situations with multiple name spaces. Knowing an author under a pen name might not mean knowing the author's birth name. Knowing students by face is very different then knowing their names or ID numbers. Children separated at birth might know each other, but not know that they are related.

Aloni refers to name spaces as "conceptual covers" (2005; 2018). In this paper, we describe the universal properties of "know-who"' that are true for any fixed name space.

Outline

The rest of the paper is organized as follows. In the next section we introduce and discuss Grove-Halpern epistemic models with names. Then, we describe syntax of our logical system and give its formal semantics. After this we highlight a possible extension of our logic by explicit names, discuss connection between our semantics and de dicto/de re knowledge, and review the related literature. In the next two sections, we introduce the axioms of the Logic of Know-Who and prove their soundness. Section Completeness Overview highlights the key steps in the proof of the completeness. The actual proof of the completeness is given in the full version of this paper (Epstein and Naumov 2020). The last section concludes.

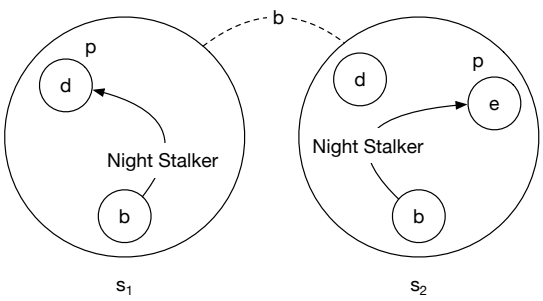

Figure 1: The Night Stalker Model (not all edges are shown). Propositional variable p means "is the murderer".

Grove-Halpern Models

The formal semantics of names that we use in this paper was first proposed by Grove and Halpern to study modality "for all agents with a given name" (1991).

Definition 1 *A tuple $(S, A, P, \{\sim_a\}_{a\in A}, N, I, \pi)$ is called a model if*

1. *S is an arbitrary set of "states",*
2. *A is an arbitrary set of "agents",*
3. *P is a function that maps each agent $a \in A$ into a set of states $P(a) \subseteq S$ in which the agent is "present",*
4. *$\sim_a$ is an "indistinguishability" equivalence relation on set $P(a)$ for each agent $a \in A$,*
5. *N is a set of "names",*
6. *$I \subseteq A \times S \times N \times A$ is an "identification mechanism" relation satisfying the following two conditions:*
   (a) *for each $a \in A$, each $s \in P(a)$, and each $n \in N$, there is at least one agent $a' \in A$ such that $(a, s, n, a') \in I$,*
   (b) *for each $a \in A$, each $s \in P(a)$, each $n \in N$, and each agent $a' \in A$, if $(a, s, n, a') \in I$, then $s \in P(a')$,*
7. *for each propositional variable p, set $\pi(p)$ is an arbitrary set of pairs (a, s) such that $a \in A$ and $s \in P(a)$.*

Figure 1 depicts a Grove-Halpern model for the Night Stalker example. Grove-Halpern models use states and indistinguishability relation $\sim_a$ to capture knowledge in almost the same way as it is done in Kripke models for the epistemic logic S5. The diagram in Figure 1 depicts two states, s_1 and s_2, indistinguishable by Birchim (b). In state s_1, DeAngelo (d) is the murderer. In state s_2, somebody else, agent e is the murderer. The significant difference between S5 models and Grove-Halpern models is that the latter do not assume that each agent is present in each state. This generalization of semantics would be insignificant in standard epistemic logic, but it is important for our logical system because its language contains modality "for all agents in the given state". To capture which agent is present in which state, in addition to the set of states S and the set of agents A, the model also includes set $P(a) \subseteq S$ for each agent $a \in A$. Set $P(a)$ is the set of states in which an agent a is "present". In the Night Stalker model, agents b and d are present in both states and agent e is only present in state s_2, see Figure 1. Thus, $P(b) = P(d) = \{s_1, s_2\}$ and $P(e) = \{s_2\}$. Intuitively, an agent cannot distinguish or not distinguish states in which she is not present. Thus, we assume that the indistinguishability relation $\sim_a$ is only defined on the set of states $P(a)$ in which the agent a is present.

As we have seen in our introductory example, the meaning of *know-who* is impossible to define without specifying the name space. Any Grove-Halpern model assumes a fixed set of names N. In our running example, $N = \{$"Night Stalker"$\}$. The identification mechanism I is the key part of defining a name space. This mechanism specifies which name could be used to refer to which agent. Grove-Halpern models take one of the most general approaches of assigning names to agents. They allow names like "my mother" that might refer to different women when used by different people. Thus, the meaning of a name is assumed to be *agent-specific*. They also allow names like "my best friend" that might refer to different people in different states. Thus, the meaning of a name is assumed to be *state-specific*. Furthermore, it is assumed that an agent might use different names to refer to the same person. Hence, for example, there could be a name space that simultaneously includes names "Night Stalker" and "East Area Rapist" for the same person. In such a name space, just like in our example, spokesman Birchim would be able to claim that he knows who the killer is even if he only can identify the perpetrator as "Night Stalker" but not as "East Area Rapist". Finally, the models allow names like "my parent" that the same person in the same state might use to refer to two different people. If one says that she knows who, her parent, raised her, then we interpret this as her saying that she knows that she was raised by both parents. To support all these features, an identification mechanism I is specified as a set of tuples $(a, s, n, a') \in A \times S \times N \times A$. If $(a, s, n, a') \in I$, then agent a in state s might use name n to refer to agent a'. The mechanism of the Night Stalker model is depicted by the directed edge on the diagram in Figure 1. For instance, the directed edge labeled with name Night Stalker from agent b to agent d inside state s_1 means that $(b, s_1, \text{Night Stalker}, d) \in I$. In other words, name Night Stalker refers to agent d when used by agent b in state s_1.

We believe that our work could be relatively easily generalized to a setting with multiple name spaces similar to one used in (Aloni 2005, 2018). If multiple name spaces would be present in the semantics, then the logical system could have multiple know-who modalities labeled by name spaces. Generally speaking, these modalities will be unrelated. In other words, knowing who in one name space does not say anything about knowing who in the other. In this paper, we restrict consideration to a single name space.

In spite of allowing very general identification mechanisms, we impose on them two restrictions captured by conditions 6(a) and 6(b) of Definition 1. The first of these conditions states that for any agent a, any state $s \in P(a)$, and any name $n \in N$, there must exist at least one agent a' that agent a refers to by name n in state s. In other words, we want to exclude cases when Birchim would claim that he knows that, say, the Santa Claus is the murderer, when there is no single person who is Santa Claus. The second condition requires

that any of the above agents a' must be present in state s. This condition guarantees that "Night Stalker" exists in the epistemic state in which Birchim knows that "Night Stalker" is the murderer. We introduce these two conditions on name spaces because we believe that without them our formal definition of "know-who" modality, see Definition 2, does not reflect the informal meaning of "knowing who".

Another important difference between Grove-Halpern models and the standard Kripke semantics for epistemic logic S5 is that valuation function π maps propositional variables not into sets of states, but into sets of pairs (a, s) consisting of an agent a and a state $s \in P(a)$ in which agent a is present. In other words, propositional variables interpreted as sentences in which the subject is omitted. Grove and Halpern call them *relative* sentences. In our example from Figure 1, the phrase "is the murderer" from the sentence "Spokesman Russ Birchim knows who is the murderer," is the meaning of proposition p. Set $\pi(p)$ is the set of all pairs (a, s) such that statement p is true about agent a in state s.

Grove and Halpern (1991) first introduce Definition 1 without conditions 6(a) and 6(b). Later they add condition 6(b) but simultaneously make names no longer agent-specific (1993). In his third work, Grove again makes names agent-specific and adds condition 6(a), but in a form stronger than ours: "exactly one" instead of "at least one" (1995). The notion of a conceptual cover (Aloni 2005) is significantly more restrictive. It requires each agent to have a unique name and each name to refer to a unique agent.

Syntax

Not only we interpret propositional variables as relative sentences, but we do the same with all modal formulae in our language. In Definition 2, we will specify formal semantics of our logic as a ternary relation $(a, s) \Vdash \varphi$ between an agent, a states, and a formula. Informally, it means that formula φ is true in state s *about* agent a. In our introductory example, $(a, s) \Vdash$ "is the murderer" where a is the person who was known as "Night Stalker". This approach allows a very straightforward treatment of know-who modality W. Namely, to state that spokesman Birchim knows who is the murderer, we write

$$(b, s) \Vdash \mathsf{W}(\text{"is the murderer"}),$$

where b is the person known as spokesman Birchim. Imagine a hypothetical situation when police announces a press conference at which Birchim will disclose the name of the murderer. Before the conference starts, any journalist j attending the conference, would not know yet who is the murderer:

$$(j, s) \Vdash \neg\mathsf{W}(\text{"is the murderer"}),$$

but the journalist would know who, spokesman Birchim, knows who is the murderer:

$$(j, s) \Vdash \mathsf{WW}(\text{"is the murderer"}).$$

We treat knowledge modality K in a similar subscript-free fashion. Namely, we write $(a, s) \Vdash \mathsf{K}\varphi$ if in a state s an agent a knows that statement φ is true *about* the agent a. For example, because "Night Stalker" knows that he himself is the murderer,

$$(a, s) \Vdash \mathsf{K}(\text{"is the murderer"}).$$

where a is the person who was known as "Night Stalker". In addition to modalities for know-who W and knowledge K, our system also includes modality A that stands for "all agents in the state". For example, the journalist would know that not all people are innocent:

$$(j, s) \Vdash \mathsf{K}\neg\mathsf{A}(\text{"is not the murderer"}).$$

Although relative sentences have already been used by Grove and Halpern (1991), subscript-free modalities were introduced much later in Friendship Logic (Seligman, Liu, and Girard 2013) that contains modalities K, A, and F. The latter stands for "for all my friends".

In this paper we propose a sound and complete logical system that describes the interplay between modalities W, K, and A. We assume a fixed countable set of propositional variables. The language Φ of our system is defined by the grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{W}\varphi \mid \mathsf{K}\varphi \mid \mathsf{A}\varphi.$$

We read $\mathsf{W}\varphi$ as "knows an agent for whom φ is true", $\mathsf{K}\varphi$ as "knows that φ is true about herself", and $\mathsf{A}\varphi$ as "φ is true for all agents". We suppose that Boolean constant $\top$ and conjunction $\wedge$ are defined in the standard way. For any finite set $X \subseteq \Phi$, by $\wedge X$ we mean the conjunction of all formulae in X. By definition, $\wedge\varnothing$ is $\top$.

Semantics

Next we define formal semantics of our logical system. The key part of this definition is item 6 that specifies the meaning of know-who modality W.

Definition 2 *For any model $(S, A, P, \{\sim_a\}_{a\in A}, N, I, \pi)$, any agent $a \in A$, any state $s \in P(a)$, and any formula $\varphi \in \Phi$, satisfiability relation $(a, s) \Vdash \varphi$ defined as follows:*

1. $(a, s) \Vdash p$ *if* $(a, s) \in \pi(p)$,
2. $(a, s) \Vdash \neg\varphi$ *if* $(a, s) \nVdash \varphi$,
3. $(a, s) \Vdash \varphi \to \psi$ *if* $(a, s) \nVdash \varphi$ *or* $(a, s) \Vdash \psi$,
4. $(a, s) \Vdash \mathsf{A}\varphi$ *if* $(a', s) \Vdash \varphi$ *for each agent* $a' \in A$ *such that* $s \in P(a')$,
5. $(a, s) \Vdash \mathsf{K}\varphi$ *if* $(a, s') \Vdash \varphi$ *for each state* $s' \in P(a)$ *such that* $s \sim_a s'$,
6. $(a, s) \Vdash \mathsf{W}\varphi$ *when there is a name* $n \in N$ *such that for each state* $s' \in P(a)$ *and each agent* $a' \in A$, *if* $s \sim_a s'$ *and* $(a, s', n, a') \in I$, *then* $(a', s') \Vdash \varphi$.

State s' in item 6 is used to capture the "know" part of "know-who". Namely, we require that the same name n identifies the right person in all states s' that agent a cannot distinguish from the current state s. This is very similar to how the modality *know-how* is often defined in the literature (Ågotnes and Alechina 2019; Fervari et al. 2017; Naumov and Tao 2017, 2018c,a,b).

In spite of its generality, our definition of know-who has limitations. Namely, it does not support the case when

"who" is a group of agents as in "John knows who is conspiring against whom" and "John knows who insulted whom in whose presence" (Boër and Lycan 2003). The complexity of these settings comes not from the fact that know-who refers to a set of agents, but rather from the fact that this set has a structure. An "insulter" is different from the "insultee" and the "observer". A hypothetical group know-who modality would need not only refer to a name of the group, but also to specify *who is who* in this group.

Explicit Names

In the standard epistemic logic, only state s is placed on the left-hand-side of the satisfiability relation $\Vdash$. As a result, statements in this logic are about states, not agents. By following (Grove and Halpern 1991, 1993; Grove 1995; Seligman, Liu, and Girard 2013) and placing both the state and the agent on the left-hand-side of $\Vdash$, we gain the ability to express statements about states, statements about agents, and statements about agents in states.

It appears, however, that we lose the ability to express statements like "in state s agent a knows that agent b knows φ", which is expressible in the standard epistemic logic by $s \Vdash \mathsf{K}_a\mathsf{K}_b\varphi$. This ability could be easily restored by adding *reference by name* modality $@_n$ to our language to form language $\Phi^@$:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{W}\varphi \mid \mathsf{K}\varphi \mid \mathsf{A}\varphi \mid @_n\varphi,$$

where n is any name. We read $@_n$ as "for any agent with name n". The semantics of language $\Phi^@$ could be defined using Grove-Halpern models by adding the following part to Definition 2:

Definition 3 *$(a, s) \Vdash @_n\varphi$ when for each agent $a' \in A$, if $(a, s, n, a') \in I$, then $(a', s) \Vdash \varphi$.*

Using modality $@$, statement "in state s agent a knows that agent b knows φ" could be written in our system as $(a, s) \Vdash \mathsf{K}@_{\text{Bob}}\mathsf{K}\varphi$, assuming that in state s agent a refers to agent b as Bob. Note that with this addition, we still retain the ability to have statements about states and agents. For example, statement φ in the above example could be any statement about state s and/or agent b. Using modality $@$ we can express the fact that the agent known to spokesman Russ Birchim as "the Night Stalker" is the murderer, see Figure 1, as

$$(b, s_1) \Vdash @_{\text{Night Stalker}}\, p.$$

We can also express the fact that Birchim knows this as

$$(b, s_1) \Vdash \mathsf{K}@_{\text{Night Stalker}}\, p.$$

In this paper we give a complete logical system that describes the universal properties expressible in language Φ, leaving proving completeness of a similar system for language $\Phi^@$ for the future.

Knowing De Dicto vs. De Re

There has been a long tradition of discussions in philosophy whether one should distinguish knowing the name of a object from knowing the object itself. These two forms of knowledge are often referred to as *de dicto* and *de re* knowledge respectively. Here is one of the examples used in the literature to distinguish these two forms of knowledge:

> *Suppose, for example, that Im asked who is Obama. While in some contexts, say at an exam at school, in order to answer it I have to know that Obama is the president of the US, in some other context, say at a party at the White House, what is needed is knowledge of someone in particular that he is Obama.* (Corsi and Orlandelli 2013)

The authors of this example consider "knowing who Obama is" in the first case as a de dicto knowledge of the fact that Obama is a name of 44th President of the United States, while "knowing who Obama is" in the second example as de re knowledge of Obama as a physical object. We disagree. To us, the only difference between these two cases is that the first is using naming system based on job title ("44th President") while the second is using naming system based on visual identity. To make our point about how artificial the distinction between knowing the name and knowing the object is, consider a hypothetical example when baby Barack Obama was accidentally switched with baby Omar Bari at birth in the hospital. As a result, Omar Bari grew up under name Barack Obama and became the 44th U.S. president, while "real" Barack Obama works as a hotel manager in Hawaii under name Omar Bari. When somebody at a party in White House is asking who is Obama, are they looking for the President or the "real" Obama, the manager?

Wang and Seligman (2018) argue for the distinction between de dicto and de re knowledge using the broken robot example originally proposed in (Grove 1995):

> *Grove gives an interesting example of a robot with a mechanical problem calling out for help (perhaps in a Matrix-like future with robots ruling the world unaided by humans). To plan further actions, the broken robot, called a, needs to know if its request has been heard by the maintenance robot, called b. But how to state exactly what a needs to know?*

To illustrate de re/de dicto distinction they list four different things that a, the broken robot, might know: (i) the robot named b knows that the robot named a needs help, (ii) the robot named b knows that it, i.e. the broken robot, needs help, (iii) the maintenance robot knows that the robot named a needs help, (iv) the maintenance robot knows that it, i.e. the broken robot, needs help. Although we agree that these four sentences have different meanings, we believe that this difference could be completely captured by distinguishing name space containing names a and b from the name space containing names "maintenance robot" and "broken robot".

Since the distinction between de dicto and de re appears unimportant in our setting, we do not stress it in this paper.

Related Literature

Hintikka (1962) argues that statement "agent a knows who is agent b" could be expressed in a first order epistemic logic as $\exists x\, \mathsf{K}_a(b = x)$. Wang agrees, stating that "to formalize 'I know who b is' we do need quantifiers" (2018a). Boër

and Lycan discuss multiple meanings of "know-who" in English (2003). Aloni adds conceptual covers (name spaces) to modal language with first-order quantifiers and proves the completeness of such system (2005). She later further develops this approach (2018). Wang and Seligman's related work, while not dealing directly with know-who, proposes a sound and complete term logic capturing properties of non-rigid names that might not be common knowledge (2018). Unlike these works, we treat know-who as a single modality and avoid the use of quantifiers.

Wang calls know-who one of "know-wh" types of knowledge: know-who, know-how, know-whether, know-what (2018a). Among them, modal properties of *know-how* are studied the most (Ågotnes and Alechina 2019; Fervari et al. 2017; Wang 2015, 2018b; Naumov and Tao 2017, 2018c,a,b; Cao and Naumov 2020). Logics of *know-whether* are studied in (Fan, Wang, and Van Ditmarsch 2015; Fan et al. 2020). Different forms of *know-value* logics are investigated in (Wang and Fan 2013; Gu and Wang 2016; van Eijck, Gattinger, and Wang 2017). Xu, Wang, and Studer proposed a logic of *know-why* (2019).

Axioms

In addition to propositional tautologies in language Φ, our logical system has the following axioms, where here and in the rest of the paper $\Box$ is either modality A or modality K:

1. Truth: $\Box\varphi \to \varphi$,
2. Distributivity: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$,
3. Negative Introspection: $\neg\Box\varphi \to \Box\neg\Box\varphi$,
4. Know-Nobody: $\mathsf{A}\neg\varphi \to \neg\mathsf{W}\varphi$,
5. Know-All: $\mathsf{KA}(\varphi \to \psi) \to (\mathsf{W}\varphi \to \mathsf{W}\psi)$,
6. Introspection of Know-Who: $\mathsf{W}\varphi \to \mathsf{KW}\varphi$.

The Truth, the Distributivity, and the Negative Introspection are standard S5 axioms. The Know-Nobody axiom says that if there is no agent in the current state for whom φ is true, then the current agent cannot know somebody for whom φ is true. The Know-All axiom says that if the agent knows that $\varphi \to \psi$ for all agents in the current state and the current agent knows someone for whom φ is true, then she also knows someone for whom ψ is true. The Introspection of Know-Who axiom says that if the current agent knows for whom φ is true, then she knows that she knows.

We write $\vdash \varphi$ if formula φ is provable in our logical system using the Modus Ponens inference rule and the three forms of the Necessitation inference rule:

$$\frac{\varphi, \quad \varphi \to \psi}{\psi} \qquad \frac{\varphi}{\mathsf{A}\varphi} \qquad \frac{\varphi}{\mathsf{K}\varphi} \qquad \frac{\varphi}{\mathsf{W}\varphi}.$$

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

The next two lemmas state well-known facts about S5 modality. We give their proofs in the full version of the paper.

Lemma 1 *If* $\varphi_1, \ldots, \varphi_n \vdash \psi$, *then* $\Box\varphi_1, \ldots, \Box\varphi_n \vdash \Box\psi$.

Lemma 2 $\vdash \Box\varphi \to \Box\Box\varphi$.

Soundness

In this section we prove the soundness of our logical system. The soundness of the Truth, the Distributivity, and the Negative Introspection axioms, as well as the Modus Ponens and the three forms of the Necessitation inference rule, is straightforward. Below we show the soundness of each remaining axiom as a separate lemma. In these lemmas we assume that (a, s) is an arbitrary pair of an agent a and a state s such that $s \in P(a)$.

Lemma 3 *If* $(a, s) \Vdash \mathsf{A}\neg\varphi$, *then* $(a, s) \nVdash \mathsf{W}\varphi$.

PROOF. Suppose that $(a, s) \Vdash \mathsf{W}\varphi$. Thus, by item 6 of Definition 2, there is a name $n \in N$ such that for each state $s' \in P(a)$ and each agent $a' \in A$, if $s \sim_a s'$ and $(a, s', n, a') \in I$, then $(a', s') \Vdash \varphi$.

Note that $s \in P(a)$ by the assumption in the preamble of this section and $s \sim_a s$ because $\sim_a$ is an equivalence relation. Thus, for each agent $a' \in A$, if $(a, s, n, a') \in I$, then $(a', s) \Vdash \varphi$.

By condition (a) of item 6 in Definition 1, there is at least one agent $a' \in A$ such that $(a, s, n, a') \in I$. Thus, $(a', s) \Vdash \varphi$. Hence, $(a', s) \nVdash \neg\varphi$ by item 2 of Definition 2. Therefore, $(a, s) \nVdash \mathsf{A}\neg\varphi$ by item 4 of Definition 2. $\boxtimes$

Lemma 4 *If* $(a, s) \Vdash \mathsf{KA}(\varphi \to \psi)$ *and* $(a, s) \Vdash \mathsf{W}\varphi$, *then* $(a, s) \Vdash \mathsf{W}\psi$.

PROOF. Suppose that $(a, s) \Vdash \mathsf{W}\varphi$. Thus, by item 6 of Definition 2, there is a name $n \in N$ such that for each state $s' \in P(a)$ and each agent $a' \in A$, if $s \sim_a s'$ and $(a, s', n, a') \in I$, then $(a', s') \Vdash \varphi$.

Consider any state $s' \in P(a)$ and any agent $a' \in A$ such that $s \sim_a s'$ and $(a, s', n, a') \in I$. Then, as we have shown above,

$$(a', s') \Vdash \varphi. \tag{1}$$

By item 6 of Definition 2, it will suffice to show that $(a', s') \Vdash \psi$. Indeed, by item 5 of Definition 2 assumption $(a, s) \Vdash \mathsf{KA}(\varphi \to \psi)$ implies that $(a, s') \Vdash \mathsf{A}(\varphi \to \psi)$ because $s' \in P(a)$ and $s \sim_a s'$.

Note that $s' \in P(a')$ by condition (b) of item 6 in Definition 1 because $(a, s', n, a') \in I$. Hence, statement $(a, s') \Vdash \mathsf{A}(\varphi \to \psi)$ implies that $(a', s') \Vdash \varphi \to \psi$ by item 4 of Definition 2. Therefore, $(a', s') \Vdash \psi$ by item 3 of Definition 2 and statement (1). $\boxtimes$

Lemma 5 *If* $(a, s) \Vdash \mathsf{W}\varphi$, *then* $(a, s) \Vdash \mathsf{KW}\varphi$.

PROOF. Consider any state $s' \in P(a)$ such that $s \sim_a s'$. By item 5 of Definition 2, it suffices to show that $(a, s') \Vdash \mathsf{W}\varphi$.

By item 6 of Definition 2, assumption $(a, s) \Vdash \mathsf{W}\varphi$ implies that there is a name $n \in N$ such that for each state $s'' \in P(a)$ and each agent $a' \in A$, if $s \sim_a s''$ and $(a, s'', n, a') \in I$, then $(a', s'') \Vdash \varphi$. Recall that $s \sim_a s'$. Thus, for each state $s'' \in P(a)$ and each agent $a' \in A$, if $s' \sim_a s''$ and $(a, s'', n, a') \in I$, then $(a', s'') \Vdash \varphi$ because $\sim_a$ is an equivalence relation. Therefore, $(a, s') \Vdash \mathsf{W}\varphi$ by item 6 of Definition 2. $\boxtimes$

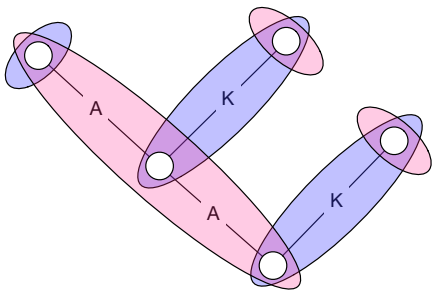

Figure 2: Nodes are views, pink A-classes are states, and blue K-classes are agents.

Completeness Overview

In this section we highlight the key steps in the proof of the completeness. The prove itself is located in the full paper.

In modal logic, a proof of a completeness usually constructs a canonical model with states being maximal consistent sets. The key property of the canonical model is normally captured by the "induction" or "truth" lemma that ordinarily states that a formula is satisfied at a state if and only if it belongs to the corresponding maximal consistent set. In our case, satisfiability is defined as a relation $(a, s) \Vdash \varphi$ between an agent a, a state s, and a formula φ. As a result, in our construction, a maximal consistent set corresponds not to a state, but to a pair (a, s) consisting of an agent a and a state s. We informally refer to such pairs as "views". The induction lemma in our paper states that a formula is satisfied at a view if and only if it belongs to the maximal consistent set corresponding to this view.

There are three distinct challenges that we faced while proving the completeness theorem. The first of them is how to define agents and states, assuming that views are maximal consistent sets of formulae. Our first attempt was based on observation that two views that have the same states satisfy exactly the same A-formulae. Thus, one can define states as classes of views (maximal consistent sets) that have the same A-formulae. Similarly, it is reasonable to assume that if two sets have exactly the same K-formulae, then they correspond to two views of the same agent in two indistinguishable states. Hence, one can define agents as classes of views that have the same K-formulae. The problem with this approach is that there could be two distinct maximal consistent sets that have the same A-formulae and the same K-formula. Such sets could be unequal because, for example, one of them contains a propositional variable and the other the negation of the same variable. Informally, such sets would correspond to two different views of the same agent in the same state. This is problematic because our formal semantics captured in Definition 2 assumes that if an agent a is present in a state s, then she has a unique view (a, s) in this state.

To solve this problem, we need to guarantee that any class of views representing a state has at most one common element with any class of views representing an agent. We achieve this by using a *tree construction*. The canonical model in our proof is a tree whose nodes are labeled with maximal consistent sets and edges are labeled with a single modality: either A or K, see Figure 2. Informally, nodes of this tree correspond to views. We say that two nodes are A-equivalent if all edges along the simple path between these two nodes are labeled with modality A and define states as equivalence classes with respect to this relation. Similarly, nodes are K-equivalent if all edges along the simple path between them are labeled with modality K. Agents are K-equivalence classes of nodes. Note that there is a unique simple path between any two nodes in a tree. As a result, the same two nodes cannot be A-equivalent and K-equivalent at the same time. Thus, this construction results in at most one node (view) corresponding to any pair consisting of an agent and a state. This guarantees that there is at most one view for any agent in any state. Of course, an agent (K-equivalence class) might have no common nodes with a state (A-equivalence class). In this case, the agent is not present in the state.

As pointed out earlier, any two views that have the same state must have the same A-formulae. We guarantee this by requiring any two nodes connected by an A-edge to have the same A-formulae. Similarly, we require any two nodes connected by a K-edge to have the same K-formulae.

Trees have previously been used in work on coalition know-how (Naumov and Tao 2017, 2018c,b,a; Cao and Naumov 2020), but for a different purpose – to model distributed knowledge. The use of trees to guarantee that intersections of classes of nodes have at most one element is an original contribution of this work.

The second major challenge that we had to overcome while proving the completeness is creating the actual nodes, or maximal consistent sets of formulae. The standard proof of completeness in modal logic usually contains a "child" lemma that for each maximal consistent set X and each formula $\neg\Box\varphi \in X$ constructs another set that contains formula $\neg\varphi$. The situation is more complicated for modality W because one needs to construct two new interdependent maximal consistent sets simultaneously: one that corresponds to view (a, s') and another to view (a', s'), see item 6 of Definition 2. Unfortunately, because of the interdependency, these two sets cannot be constructed consecutively. To construct them simultaneously, we developed a new technique that consists in defining a property of a pair of sets of formulae, choosing a pair of small sets satisfying this property, and then extending the sets while maintaining the property. When fully extended, each of the sets will become the label of a node in the tree construction that we described above and will represent a view in our model. Informally, the property that we maintain could be described as "views can coexist in the same states". We call such views *consonant*. A somewhat similar construction of two interdependent nodes has been used in (Naumov and Tao 2018c,a) to construct two states of a game in "harmony". The construction proposed in this paper creates two nodes that belong to the same state and, thus have the same A-formulae. The two states in "harmony" are consecutive states of a game that do not share any specific class of formulae. As a result, the properties of

consonant pairs are different from properties of pairs in "harmony" and the proofs that the corresponding constructions work are also different.

The third challenge in constructing the canonical model is to define the right identification mechanism. What name should one of the blue classes (agents) in Figure 2 use to refer to another blue class in one of the pink classes (states)? The solution that we propose at first sounds unbelievably simple. In essence, when spokesman Birchim knows who is the killer, we want phrase "is the killer" to be the name under which Birchim knows the killer. In general if $(a, s) \Vdash \mathsf{W}\varphi$, then formula φ itself is the name under which agent a knows the agent with property φ in state s. Although elegant, this naming scheme has a fatal flaw: it does not distinguish between knowing that an agent *exists* and *knowing who* the agent is. For example, using this identification mechanism, spokesman Birchim would know who is the killer (agent named "is the killer") the moment Birchim is notified that the murder is committed. Similarly, a journalist arriving to the press conference would know who is the killer even before the conference starts. In general, this naming space makes formula $\mathsf{K}\neg\mathsf{A}\neg\varphi \rightarrow \mathsf{W}\varphi$ true in any model that uses this identification mechanism. Since this formulae is not universally valid, the mechanism cannot be used in the canonical model construction of the completeness proof.

We solve this problem by modifying the above identification mechanism. We still allow "is the killer" as the name, but we say that, when used by spokesman Birchim, this name refers to the actual killer only if in the current state Birchman actually knows who the killer is. Otherwise, when used by him, this name refers to all agents present in the state. Thus, if Birchman knows who the killer is, then he can use name "is the killer" to identify the killer, otherwise, he cannot. We are now ready to answer our prior question regarding names used by blue classes (agent) at pink classes (states) in Figure 2. If the maximal consistent set of unique node at the intersection of an agent a and a state s contains formula $\mathsf{W}\varphi$, then name φ, when used by agent a at state s, refers to all agents b present in the state s such that the maximal consistent set of the unique node at the intersection of agent b and state s contains formula φ. Otherwise, name φ refers to all agents present in state s.

As an example, consider the fragment of the tree depicted in Figure 3. Nodes u and t are connected by an A-edge. Thus, they represent views of two different agents, a_1 and a_2, in the same state s_1. On the other hand, nodes u and v are connected by a K-edge. Hence, they represent two views of the same agent a_2 in two different (but indistinguishable to the agent) states: s_1 and s_2. Note that agent a_1 is not present in state s_2 because the corresponding ovals have no common nodes. The maximal consistent sets associated with views u and v contain formula $\mathsf{W}p$. As a result, when name p is used in these two views, it refers to the agents in the same state whose maximal consistent sets contain variable p. In other words, when name p is used by agent a_2 in state s_1, it refers to agent a_1 and when the same name is used by the same agent in state s_2, it refers to agent a_2 herself. At the same time, because formula $\mathsf{W}q$ does not belong to the maximal consistent sets corresponding to nodes u and v, when name

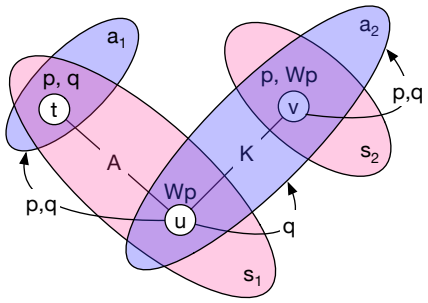

Figure 3: Fragment of a Canonical Model.

q is used in these two views, it refers to all agents present in the corresponding state. In other words, in state s_1 name q is used by agent a_2 to refer to herself and agent a_1; in state s_2 the same name is used by the same agent to refer only to herself.

This concludes the overview of the proof of the strong completeness theorem stated below. The complete proof can be found in the full paper.

Theorem 1 *If $X \nvdash \varphi$, then there is an agent $a \in A$ and a state $s \in P(a)$ of a model $(S, A, P, \{\sim_a\}_{a\in A}, N, I, \pi)$ such that $(a, s) \Vdash \chi$ for each formula $\chi \in X$ and $(a, s) \nVdash \varphi$.*

Conclusion

The contribution of this paper is three-fold. First, we proposed a formal semantics of modality *know-who* which is based on Grove-Halpern epistemic models with names. Second, following (Seligman, Liu, and Girard 2013) we propose a syntax for this modality that does not require the use of agent subscript. Without this modification to the language it would be hard to express statements like "a journalist knows who knows who the murderer is". Finally, we give a complete logical system that describes the interplay between modalities "know-who", "know", and "for all agents". We believe that the standard filtration technique from modal logic could be used to prove weak completeness of our logical system with respect to the class of finite models. This would imply that our system, unlike logics with quantifiers previously used to capture know-who, is decidable. We also believe that the results in this paper could be generalized to a logical system that supports multiple name spaces. Each such name space η will have its own identification mechanism I_η in Definition 1 and its own modality W_η.

References

Ågotnes, T., and Alechina, N. 2019. Coalition logic with individual, distributed and common knowledge. *Journal of Logic and Computation* 29:1041–1069.

Aloni, M. 2005. Individual concepts in modal predicate logic. *Journal of Philosophical Logic* 34(1):1–64.

Aloni, M. 2018. Knowing-who in quantified epistemic logic. In *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*. Springer. 109–129.

Boër, S. E., and Lycan, W. G. 2003. *Knowing Who*. MIT Press.

Cao, R., and Naumov, P. 2020. Knowing the price of success. *Artificial Intelligence* 103287.

Corsi, G., and Orlandelli, E. 2013. Free quantified epistemic logics. *Studia Logica* 101(6):1159–1183.

Epstein, S., and Naumov, P. 2020. Epistemic logic of know-who. *arXiv:2012.06651*.

Fan, J.; Grossi, D.; Kooi, B.; Su, X.; and Verbrugge, R. 2020. Commonly knowingly whether. *arXiv:2001.03945*.

Fan, J.; Wang, Y.; and Van Ditmarsch, H. 2015. Contingency and knowing whether. *The Review of Symbolic Logic* 8(1):75–107.

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.

Grove, A. J., and Halpern, J. Y. 1991. Naming and identity in a multi-agent epistemic logic. *KR* 91:301–312.

Grove, A. J., and Halpern, J. Y. 1993. Naming and identity in epistemic logics part i: the propositional case. *Journal of Logic and Computation* 3(4):345–378.

Grove, A. J. 1995. Naming and identity in epistemic logic part ii: a first-order logic for naming. *Artificial Intelligence* 74(2):311–350.

Gu, T., and Wang, Y. 2016. "Knowing value" logic as a normal modal logic. In Lev Beklemishev, S. D., and Máté, A., eds., *Advances in Modal Logic 11, proceedings of the 11th conference on "Advances in Modal Logic," held in Budapest, Hungary, from 30 August to 2 September 2016*, 362–381. College Publications.

Hardy, D. 1981. Two found slain in goleta; case similar to one in '79. *Santa Barbara News-Press*. July 28th, http://www.goldenstatekiller.com/1981-07-28.pdf.

Hintikka, J. 1962. *Knowledge and Belief - An Introduction to the Logic of the Two Notions*. Contemporary philosophy. Ithaca, NY: Cornell University Press.

Hurst, J. 1981. "Night Stalker" theory connecting eight southland slayings disputed. *Los Angeles Times* p.3. August 2nd, https://www.newspapers.com/clip/19563813/night_stalker_theory_connecting_eight/, last accessed February 28, 2021.

Levine, N. 2018. Read the east area rapist stories that gripped sacramento in 1977. *The Saacramento Bee*. April 26th, https://www.sacbee.com /news/local/crime/article209913329.html, last accessed February 28, 2021.

Naumov, P., and Tao, J. 2017. 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.

Seligman, J.; Liu, F.; and Girard, P. 2013. Facebook and the epistemic logic of friendship. In *14th conference on Theoretical Aspects of Rationality and Knowledge (TARK '13), January 2013, Chennai, India*, 229–238.

Serna, J., and Oreskes, B. 2018. Must Reads: Why did it take so long to arrest the Golden State Killer suspect? Interagency rivalries, old technology, errors and bad luck. *Los Angeles Times*. , May 25th, https://www.latimes.com/local/lanow/la-me-ln-golden-state-killer-case-20180525-story.html, last accessed February 28, 2021.

van Eijck, J.; Gattinger, M.; and Wang, Y. 2017. Knowing values and public inspection. In *Indian Conference on Logic and Its Applications*, 77–90. Springer.

Wang, Y., and Fan, J. 2013. Knowing that, knowing what, and public communication: Public announcement logic with Kv operators. In *Twenty-Third International Joint Conference on Artificial Intelligence*.

Wang, Y., and Seligman, J. 2018. When names are not commonly known: Epistemic logic with assignments. In Bezhanishvili, G.; D'Agostino, G.; Metcalfe, G.; and Studer, T., eds., *Advances in Modal Logic 12, proceedings of the 12th conference on "Advances in Modal Logic," held in Bern, Switzerland, August 27-31, 2018*, 611–628. College Publications.

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

Wang, Y. 2018a. Beyond knowing that: A new generation of epistemic logics. In *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*. Springer. 499–533.

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

Xu, C.; Wang, Y.; and Studer, T. 2019. A logic of knowing why. *Synthese* 1–27.