A New Hope

Hans van Ditmarsch

Open University of the Netherlands

Krisztina Fruzsa [1] and Roman Kuznets [2]

TU Wien, Austria

Abstract

Knowledge has long been identified as an inherent component of agents' decision-making in distributed systems. However, for agents in fault-tolerant distributed systems with fully byzantine agents, achieving knowledge is, in most cases, unrealistic. If agents can both lie and themselves be mistaken, then a message received is generally not sufficient to create knowledge. This problem is adequately addressed by an epistemic modality named *hope*, which has already been axiomatized. In this paper, we propose an alternative complete axiomatization for the hope modality by removing the reliance on designated atoms denoting correctness of individual agents and show that hope can be described as a $\mathsf{KB4}_n$ system. This additionally brings a more streamlined presentation of the common hope modality (the hope analog of common knowledge). We also combine $\mathsf{KB4}_n$ hope modalities with $\mathsf{S5}_n$ knowledge modalities traditionally used in the epistemic analysis of fault-free distributed systems and present a logic enriched with both common knowledge and common hope. In these logics we formalize as frame-characterizable axioms some of the main properties of fully byzantine distributed systems: bounds on the number of faulty agents and the epistemic limitations due to agents' inability to rule out brain-in-a-vat scenarios.

Keywords: Epistemic logic, distributed systems, byzantine agents.

1 Introduction

Over at least three decades, epistemic analysis has been used as a potent tool [7, 14] for studying distributed systems. It is often based on the *runs and systems framework* that views global states of a distributed system as possible worlds in a Kripke model. The importance of this methodology is underscored by the broadly applicable *Knowledge of Preconditions Principle* [21], recently formulated by Moses, which states that in all models of distributed systems, if φ is a necessary condition for agent i to perform an action, then agent i knowing φ to hold, written $K_i\varphi$, is also a necessary condition for this agent

[1] PhD student in the Austrian Science Fund (FWF) doctoral program LogiCS (W1255).

[2] Funded by the FWF ByzDEL project (P33600).

to perform this action. The agent's complete reliance on its local state as the source of information about the system naturally induces an equivalence relation on the global states, resulting in agents' knowledge being described using multimodal epistemic logic $\mathsf{S5}_n$.

This epistemic analysis via the runs and systems framework was recently [15–17] extended to *fault-tolerant systems* with so-called *byzantine* agents [18]. (Fully) byzantine agents are the worst-case faulty agents to participate in a distributed system: not only can they arbitrarily deviate from their respective protocols, but their perception of their own actions and the events they observe can be corrupted, possibly unbeknownst to them, resulting in false memories. Whether byzantine agents are actually present in a system or not, the very possibility of their presence has drastic and debilitating effects on the epistemic state of all agents, due to their inability to rule out so-called *brain-in-a-vat* scenarios [23]. In a distributed system, a brain-in-a-vat agent is a faulty agent with completely corrupted perceptions that provide no reliable information about the system [16]. It has been shown that agents' inability to rule out being a brain in a vat precludes them from knowing many basic facts, including their own correctness/faultiness, in both asynchronous [16] and synchronous [25] distributed systems.

The extended runs and systems framework was used in [11] to analyze the *Firing Rebels with Relay* (FRR) problem, a simplified version of the *consistent broadcasting* primitive [26], which has been used as a pivotal building block in distributed algorithms, e.g., for byzantine fault-tolerant clock synchronization [5, 12, 24, 26, 28], synchronous consensus [27], and a general reduction of distributed task solvability in byzantine systems to solvability in systems with crash failures [19]. Instead of knowledge (unattainable due to brain-in-a-vat scenarios), the analysis of FRR hinges on a weaker epistemic notion called *hope*, which, in the presence of knowledge modalities, was initially defined as $H_i\varphi := \mathit{correct}_i \to K_i(\mathit{correct}_i \to \varphi)$. Without knowledge, hope was axiomatized in [10] with the help of designated atoms $\mathit{correct}_i$ representing agent i's correctness. The special nature of these atoms precluded the logic from being a normal modal logic.

Contributions and paper organization: In this paper, we provide an alternative axiomatization of hope that deals away with these atoms treating them as abbreviations $\mathit{correct}_i := \neg H_i\bot$ instead. Not only does this make the logic of hope a normal modal logic, but it turns out to coincide with multimodal $\mathsf{KB4}_n$ (Sect. 2). We explore the language with both hope and knowledge modalities by formulating a combined logic of hope and knowledge including their interaction and showing the Kripke completeness (Sect. 3). We also demonstrate the utility of this logic by providing frame-characterizable axioms to represent various properties of fully byzantine agents, including brain-in-a-vat scenarios, and system specifications, e.g., the upper bound on the number of faulty agents (Sect. 4). Working towards the epistemic analysis of group actions, we axiomatize the logic of common hope and common knowledge and provide the completeness theorem (Sect. 5). In Sect. 6, we take an in-depth look

P : all propositional tautologies

K^H:	$H_i(\varphi \to \psi) \land H_i\varphi \to H_i\psi$	T'^H :	$correct_i \to (H_i\varphi \to \varphi)$
4^H :	$H_i\varphi \to H_iH_i\varphi$	F :	$faulty_i \to H_i\varphi$
5^H :	$\neg H_i\varphi \to H_i\neg H_i\varphi$	H :	$H_i correct_i$
MP:	$\dfrac{\varphi \quad \varphi \to \psi}{\psi}$	Nec^H:	$\dfrac{\varphi}{H_i\varphi}$

Fig. 1. Axiom system $\mathscr{H}_{\text{co}}$ in language $\mathcal{L}_H^{\text{co}} = \mathcal{L}(\mathsf{Prop} \sqcup \mathsf{Co}, H_1, \dots, H_n)$ from [10]

at the related work. Finally, in Sect. 7, we provide conclusions and directions for future work.

2 Logic of Individual Hope

We fix a finite set $\mathcal{A} = \{1, \dots, n\}$ of *agents*, a countably infinite set Prop of *atomic propositions* (*atoms*), and a finite set $\mathsf{Co} := \{correct_i \mid i \in \mathcal{A}\}$ of designated *correctness atoms* such that $\mathsf{Prop} \cap \mathsf{Co} = \varnothing$. We will consider a number of multimodal languages varying in modalities and atoms. Hence, it pays off to give a general definition.

Definition 2.1 A *multimodal language* $\mathcal{L}(P, \heartsuit_1, \dots, \heartsuit_m)$ with a set P of atoms and with modalities $\heartsuit_j$ is defined according to the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \land \varphi) \mid \heartsuit_j\varphi$$

where $p \in P$ and $j = 1, \dots, m$. We take $\top$ to be an abbreviation for some fixed propositional tautology, $\bot := \neg\top$ and use standard abbreviations for the remaining boolean connectives.

We first consider language $\mathcal{L}_H^{\text{co}} := \mathcal{L}(\mathsf{Prop} \sqcup \mathsf{Co}, H_1, \dots, H_n)$ for *hope modalities* H_i with n designated atoms $correct_i$, one for each $i \in \mathcal{A}$, intended to mean that agent i has not deviated from its normative behavior in the system. For instance, in Kripke models generated based on runs of a fault-tolerant distributed system with perfect recall, $correct_i$ should signify that, by the time of evaluation, agent i has not violated its protocol and has correctly recorded its performed actions and witnessed events. We abbreviate $faulty_i := \neg correct_i$.

Axiom system $\mathscr{H}_{\text{co}}$ is presented in Fig. 1. Here axioms P, K^H, 4^H, and 5^H, along with rules MP and Nec^H represent standard multimodal logic $\mathsf{K45}_n$, in particular, postulating *positive* and *negative introspection* for the hope modality. Axiom T'^H is *factivity* restricted to correct agents. Axioms H and F represent further properties of $correct_i$: namely, that agents always hope to be correct and that the hopes of faulty agents are unrestricted and all encompassing, in particular, alongside tautologies they also hope for contradictions, making their hopes inconsistent. Intuitively, these properties mean that an agent can rely on its perceptions iff the agent is correct. We elaborate more on the origins of this particular axiomatization in the next section, where hope is related to the knowledge modality.

Definition 2.2 A *Kripke frame* for a language $\mathcal{L} = \mathcal{L}(P, \heartsuit_1, \dots, \heartsuit_m)$ is a tuple $F = (W, R_1, \dots, R_m)$ where $W \neq \varnothing$ is the set of *worlds* (or *states*)

$$P: \quad \text{all propositional tautologies}$$

$$K^H: \quad H_i(\varphi \to \psi) \wedge H_i\varphi \to H_i\psi \qquad B^H \ : \ \varphi \to H_i\neg H_i\neg\varphi$$

$$4^H \ : \ H_i\varphi \to H_iH_i\varphi$$

$$MP: \ \frac{\varphi \quad \varphi \to \psi}{\psi} \qquad Nec^H: \ \frac{\varphi}{H_i\varphi}$$

Fig. 2. Axiom system $\mathscr{H}$ in language $\mathcal{L}_H = \mathcal{L}(\mathsf{Prop}, H_1, \ldots, H_n)$

and $R_j \subseteq W \times W$ is an *accessibility relation* for the modality $\heartsuit_j$. A *Kripke model* for $\mathcal{L}$ is $M = (F, \pi)$ where F is a Kripke frame and $\pi\colon P \to \mathcal{P}(W)$ is a *valuation function*. Truth for formulas $\varphi \in \mathcal{L}$ in model M is defined as follows: $M, w \models p$ iff $w \in \pi(p)$ for all $p \in P$, negation and conjunction behave classically within each world, and $M, w \models \heartsuit_j\varphi$ iff $M, v \models \varphi$ for all $v \in R_j(w)$ where $R_j(w) := \{w' \mid wR_jw'\}$. *Validity in model* M, denoted $M \models \varphi$, is defined as truth in all worlds of W. *Validity in frame* F, denoted $F \models \varphi$, is defined as validity in all models (F, π). *Validity in a class* $\mathcal{C}$ *of Kripke frames* (*models*) is defined as validity in all frames (models) of $\mathcal{C}$.

A binary relation R_j is called

- *transitive* if wR_jv whenever wR_ju and uR_jv,
- *symmetric* if wR_jv whenever vR_jw,
- *euclidean* if wR_jv whenever uR_jw and uR_jv, and
- *shift serial* if $R_j(v) \neq \varnothing$ for any $v \in R_j(w)$.

Class $\mathcal{K}\mathcal{4}5_m$ ($\mathcal{K}\mathcal{B}\mathcal{4}_m$; $\mathcal{S}5_m$) consists of all models with m transitive and euclidean (transitive and symmetric; equivalence) accessibility relations. A *partial equivalence relation* is any transitive and symmetric binary relation (see [20]).

From now till the end of this section, we set $m = n = |\mathcal{A}|$ and use $R_i = \mathcal{H}_i$ as the accessibility relation for modality $\heartsuit_i = H_i$.

Definition 2.3 *Class* $\mathcal{K}\mathcal{4}5_n^{\mathrm{co}}$ consists of all Kripke models $M = \big((W, \mathcal{H}_1, \ldots, \mathcal{H}_n), \pi\big) \in \mathcal{K}\mathcal{4}5_n$ such that for every $i \in \mathcal{A}$ and $w, w' \in W$: (i) if $w \in \pi(\mathit{correct}_i)$, then $w\mathcal{H}_iw$; (ii) if $w \notin \pi(\mathit{correct}_i)$, then $\mathcal{H}_i(w) = \varnothing$; and (iii) if $w\mathcal{H}_iw'$, then $w' \in \pi(\mathit{correct}_i)$.

Theorem 2.4 ([10]) *$\mathscr{H}_{\mathrm{co}}$ is sound and complete with respect to $\mathcal{K}\mathcal{4}5_n^{\mathrm{co}}$.*

Note that class $\mathcal{K}\mathcal{4}5_n^{\mathrm{co}}$ of models is not based on any class of frames. Our first result in this paper is an alternative axiomatization for hope that deals away with designated atoms $\mathit{correct}_i$ and, hence, enables us to avoid the dependency of accessibility relations $\mathcal{H}_i$ on valuation function π. This is achieved by adopting the definition

$$\mathit{correct}_i := \neg H_i\bot \tag{1}$$

in language $\mathcal{L}_H := \mathcal{L}(\mathsf{Prop}, H_1, \ldots, H_n)$ based on the view that faulty_i can be equated to the inconsistent hopes $H_i\bot$. It turns out that the logic of hope in this language is the logic of class $\mathcal{K}\mathcal{B}\mathcal{4}_n$ of all transitive and symmetric frames and is axiomatized by axiom system $\mathscr{H} = \mathsf{KB4}_n$ presented in Fig. 2. It is well known that

Theorem 2.5 (Completeness for the logic of hope)
$\mathscr{H}$ is sound and complete with respect to $\mathcal{KB}4_n$.

Remark 2.6 The new axiomatization makes it easier to see how hope is different from the notion of belief. Indeed, belief is quite often assumed to be consistent, i.e., satisfying axiom $\neg B\bot$, which fails for hope due to inconsistent hopes of faulty agents. On the other hand, axiom B^H is typically invalid for belief because, together with 4^H, it would preclude agents from having consistent but false beliefs.

We now show that $\mathscr{H}$ is equivalent to $\mathscr{H}_{\text{co}}$ modulo abbreviation (1):

Lemma 2.7
- *$\mathscr{H} \vdash \varphi$ implies $\mathscr{H}_{\text{co}} \vdash \varphi$ for any $\varphi \in \mathcal{L}_H$.*
- *$\mathscr{H}_{\text{co}} \vdash \varphi$ implies $\mathscr{H} \vdash \varphi^\dagger$, where $\varphi^\dagger \in \mathcal{L}_H$ is the result of replacing each $correct_i$ in $\varphi \in \mathcal{L}_H^{\text{co}}$ with $\neg H_i\bot$, according to* (1).

Proof.
- It is sufficient to show $\mathscr{H}_{\text{co}} \vdash B^H$. Using the instance $faulty_i \to H_i\neg H_i\neg\varphi$ of F, by prop. reasoning, $\mathscr{H}_{\text{co}} \vdash faulty_i \to (\varphi \to H_i\neg H_i\neg\varphi)$. On the other hand, from the instance $correct_i \to (H_i\neg\varphi \to \neg\varphi)$ of T'^H, by prop. reasoning $\mathscr{H}_{\text{co}} \vdash correct_i \to (\varphi \to \neg H_i\neg\varphi)$. Coupling this with the instance $\neg H_i\neg\varphi \to H_i\neg H_i\neg\varphi$ of 5^H, we get $\mathscr{H}_{\text{co}} \vdash correct_i \to (\varphi \to H_i\neg H_i\neg\varphi)$. Since $faulty_i \vee correct_i$ is a propositional tautology, $\mathscr{H}_{\text{co}} \vdash \varphi \to H_i\neg H_i\neg\varphi$ by prop. reasoning.
- It is sufficient to show that axiom 5^H, as well as the †-translations of axioms T'^H, F, and H are derivable in $\mathscr{H}$. That 5^H can be derived from 4^H and B^H is a well-known fact (any transitive and symmetric relation is euclidean). Thus, we only discuss the axioms involving $correct_i$.
  - The †-translation of T'^H is $\neg H_i\bot \to (H_i\psi \to \psi)$ for $\psi = \varphi^\dagger$. It is sufficient to show the derivability $\mathscr{H} \vdash \neg(H_i\psi \to \psi) \to H_i\bot$ for the contrapositive. Firstly, $\mathscr{H} \vdash \neg(H_i\psi \to \psi) \to H_i\psi \wedge \neg\psi$ is a propositional tautology. Further, $\mathscr{H} \vdash H_i\psi \to H_iH_i\psi$ by 4^H and also $\mathscr{H} \vdash \neg\psi \to H_i\neg H_i\psi$ by B^H. Thus, combining these together, $\mathscr{H} \vdash \neg(H_i\psi \to \psi) \to H_iH_i\psi \wedge H_i\neg H_i\psi$. It remains to use the normality of H_i and prop. reasoning to replace $H_iH_i\psi \wedge H_i\neg H_i\psi$ first with $H_i(H_i\psi \wedge \neg H_i\psi)$ and finally with $H_i\bot$.
  - The †-translation of F is (modulo a double negation) $H_i\bot \to H_i\psi$, which follows by the normality of H_i from $\bot \to \psi$.
  - The †-translation of axiom H is $H_i\neg H_i\bot$, which is easy to obtain from the instance $\neg\bot \to H_i\neg H_i\neg\neg\bot$ of B^H by prop. reasoning. □

Theorem 2.8 (Equivalence of the two logics of individual hope)
Systems $\mathscr{H}$ and $\mathscr{H}_{\text{co}}$ are equivalent representations of the logic of hope.

We demonstrate the utility of this reformulation of the logic of hope by encoding a standard limitation on the number of faulty agents in a fault-tolerant distributed system as a *frame-characterizable* property in logic $\mathscr{H}$. It is typical to design distributed protocols under the assumption that no more than f of the n agents can become faulty ($0 \leq f < n$). This is a natural restriction

given that clearly no outcome of agents' protocols can be guaranteed if, e.g., all agents can ignore these protocols. Moreover, byzantine consensus [9, 18] and byzantine clock synchronization [6,9], among others, are unsolvable for $n \leq 3f$, while, e.g., consensus in asynchronous systems with the weakest failure detector Omega is unsolvable already for $n \leq 2f$ [4]. We can encode such requirements in $\mathcal{L}_H$ by an additional axiom

$$Byz_f \qquad := \qquad \bigvee_{\substack{G\subseteq\mathcal{A}\\ |G|=n-f}} \bigwedge_{i\in G} \neg H_i\bot.$$

Remark 2.9 $Byz_0 = \bigwedge_{i\in\mathcal{A}} \neg H_i\bot$ simply states that all n agents are correct.

Proposition 2.10 (Frame characterization of $\leq f$ faulty agents) *Axiom Byz_f is characterized by the property of frames $F = (W, \mathcal{H}_1, \ldots, \mathcal{H}_n)$*

$$(\forall w \in W)(\exists G \subseteq \mathcal{A})\Big(|G| = n - f \wedge (\forall i \in G)\mathcal{H}_i(w) \neq \varnothing\Big),$$

which we call all-but-f-seriality*. In other words, each world must have outgoing arrows for all but f agents.*

Proof. Take an arbitrary frame $F = (W, \mathcal{H}_1, \ldots, \mathcal{H}_n)$ for language $\mathcal{L}_H$. We need to show that

$$F \models Byz_f \qquad \Longleftrightarrow \qquad F \text{ is all-but-}f\text{-serial.}$$

We prove the ($\Longrightarrow$) direction by contrapositive. If F is not all-but-f-serial, there is some $w \in W$ such that any group $G \subseteq \mathcal{A}$ of $n-f$ agents has some agent $i_G \in G$ such that $\mathcal{H}_{i_G}(w) = \varnothing$. Since for all these agents, $(F, \pi), w \not\models \neg H_{i_G}\bot$ for any π, we have $(F, \pi), w \not\models Byz_f$ and, hence, $F \not\models Byz_f$.

For the ($\Longleftarrow$) direction, let F be all-but-f-serial. Take an arbitrary $w \in W$. It now follows that there is a group $G \subseteq \mathcal{A}$ of $n-f$ agents such that $\mathcal{H}_i(w) \neq \varnothing$ for all $i \in G$. Therefore, $(F, \pi), w \models \bigwedge_{i\in G} \neg H_i\bot$ and $(F, \pi), w \models Byz_f$ for any π. The validity in F follows because w and π were chosen arbitrarily. □

Definition 2.11 *Class $\mathcal{KB}4_n^{n-f}$* consists of all models from $\mathcal{KB}4_n$ with all-but-f-serial frames.

Corollary 2.12 *$\mathscr{H} + Byz_f$ is sound and complete with respect to $\mathcal{KB}4_n^{n-f}$.*

3 Logic of Individual Hope and Individual Knowledge

In this section, we consider language $\mathcal{L}_{KH} := \mathcal{L}(\mathsf{Prop}, K_1, \ldots, K_n, H_1, \ldots, H_n)$. This language with both hope and knowledge modalities for each agent is expressive enough to describe most of the epistemic attitudes relevant to distributed systems and explore relationships among them. Accordingly, the semantics used until the end of the paper has $m = 2n = 2|\mathcal{A}|$ accessibility relations where, in addition to $\mathcal{H}_i$ for hope modalities H_i we now use accessibility relations $\mathcal{K}_i$ for knowledge modalities K_i.

In particular, we now recall how hope was initially defined via knowledge and the correctness atoms. The hope modality appeared in the analysis of the Firing Rebels Problem [11], as well as in earlier works [10, 16], in the form of derived modality $H_i\varphi := \mathit{correct}_i \to K_i(\mathit{correct}_i \to \varphi)$, which, translated into language $\mathcal{L}_{KH}$ using (1), becomes axiom

$$KH \qquad := \qquad H_i\varphi \quad \leftrightarrow \quad \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big). \tag{2}$$

Our new language with hope enables us to (almost) characterize this connection axiom by two frame properties for two directions of equivalence (2).

Proposition 3.1 (Characterizing knowledge-to-hope connection) *On the class of frames $F = (W, \mathcal{K}_1, \dots, \mathcal{K}_n, \mathcal{H}_1, \dots, \mathcal{H}_n)$ with shift serial $\mathcal{H}_i$,*

$$KH^{\leftarrow} \qquad := \qquad \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big) \to H_i\varphi \tag{3}$$

is characterized by frame property

$$\mathcal{H}\text{in}\mathcal{K}: \qquad \mathcal{H}_i \subseteq \mathcal{K}_i.$$

Proof. First assume that a frame F with shift serial $\mathcal{H}_i$ satisfies $\mathcal{H}\text{in}\mathcal{K}$ and let $M = (F, \pi)$ for an arbitrary π. Let the antecedent of (3) hold at an arbitrary world $w \in W$. To show that $M, w \models H_i\varphi$, it is sufficient to show that $M, v \models \varphi$ for all $v \in \mathcal{H}_i(w)$. It is vacuously true if $\mathcal{H}_i(w) = \varnothing$. Otherwise, take any such world v. $M, w \models \neg H_i\bot$ because $\mathcal{H}_i(w) \neq \varnothing$, thus, $M, w \models K_i(\neg H_i\bot \to \varphi)$. Since $\mathcal{H}_i(w) \subseteq \mathcal{K}_i(w)$ due to $\mathcal{H}\text{in}\mathcal{K}$, we get $M, v \models \neg H_i\bot \to \varphi$. But $\mathcal{H}_i(v) \neq \varnothing$ due to the shift seriality of $\mathcal{H}_i$. This is sufficient to conclude that $M, v \models \varphi$, completing the proof that $KH^{\leftarrow}$ is valid in F.

For the opposite direction, assume that F violates $\mathcal{H}\text{in}\mathcal{K}$, i.e., that there are worlds $w, v \in W$ with $w\mathcal{H}_i v$ but not $w\mathcal{K}_i v$. Consider any model $M = (F, \pi)$ with a valuation π such that $\pi(p) = W \setminus \{v\}$ for some atom p. We have $M, w \models K_i(\neg H_i\bot \to p)$ because $\mathcal{K}_i(w) \subseteq W \setminus \{v\} = \pi(p)$. Therefore, $M, w \models \neg H_i\bot \to K_i(\neg H_i\bot \to p)$. However, clearly $M, w \not\models H_i p$ because of v. Thus, we have shown that $M, w \not\models KH^{\leftarrow}$ for $\varphi = p$. Note that this direction does not rely on the shift seriality of $\mathcal{H}_i$. □

Proposition 3.2 (Characterizing hope-to-knowledge connection)

$$KH^{\rightarrow} \qquad := \qquad H_i\varphi \to \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big) \tag{4}$$

for frames $F = (W, \mathcal{K}_1, \dots, \mathcal{K}_n, \mathcal{H}_1, \dots, \mathcal{H}_n)$ is characterized by property

$$\text{one}\mathcal{H}: \qquad (\forall w, v \in W)\big(\mathcal{H}_i(w) \neq \varnothing \wedge \mathcal{H}_i(v) \neq \varnothing \wedge w\mathcal{K}_i v \Longrightarrow w\mathcal{H}_i v\big).$$

Proof. First assume that F satisfies one$\mathcal{H}$ and let $M = (F, \pi)$ for an arbitrary π. Let $M, w \models H_i\varphi$. The case of $\mathcal{H}_i(w) = \varnothing$ is trivial since $M, w \models H_i\bot$ makes the succedent of (4) true at w. Otherwise, $\mathcal{H}_i(w) \neq \varnothing$. Similarly, for any $v \in \mathcal{K}_i(w)$ with $\mathcal{H}_i(v) = \varnothing$, we have $M, v \models \neg H_i\bot \to \varphi$. Finally, for any $v \in \mathcal{K}_i(w)$ with $\mathcal{H}_i(v) \neq \varnothing$, we have $v \in \mathcal{H}_i(w)$ by one$\mathcal{H}$. Hence, $M, v \models \varphi$

$$
\begin{array}{lll}
 & P: & \text{all propositional tautologies} \\
H^{\dagger}: & H_i\neg H_i\bot & K^K: \quad K_i(\varphi\to\psi)\wedge K_i\varphi\to K_i\psi \\
 & & 4^K: \quad K_i\varphi\to K_iK_i\varphi \\
 & & 5^K: \quad \neg K_i\varphi\to K_i\neg K_i\varphi \\
 & & T^K: \quad K_i\varphi\to\varphi \\
MP: & \dfrac{\varphi \quad \varphi\to\psi}{\psi} & Nec^K: \ \dfrac{\varphi}{K_i\varphi} \\
 & KH: \quad H_i\varphi\leftrightarrow\big(\neg H_i\bot\to K_i(\neg H_i\bot\to\varphi)\big) &
\end{array}
$$

Fig. 3. Axiom system $\mathscr{KH}$ in language $\mathcal{L}_{KH}=\mathcal{L}(\mathsf{Prop},K_1,\dots,K_n,H_1,\dots,H_n)$

and $M,v\models\neg H_i\bot\to\varphi$. We have shown that $\neg H_i\bot\to\varphi$ is true in all worlds from $\mathcal{K}_i(w)$ and can again conclude that the succedent of (4) is true at w. This completes the proof that $KH^{\to}$ is valid in F.

For the opposite direction, assume that F violates one$\mathcal{H}$, i.e., there are worlds $w,v\in W$ with $\mathcal{H}_i(w)\neq\varnothing$, $\mathcal{H}_i(v)\neq\varnothing$, $w\mathcal{K}_iv$, but not $w\mathcal{H}_iv$. Consider any model $M=(F,\pi)$ with a valuation π such that $\pi(p)=\mathcal{H}_i(w)$ for some atom p. Clearly, $M,w\models H_ip$ and $M,w\models\neg H_i\bot$. However, $M,v\models\neg H_i\bot$ and $M,v\not\models p$. Hence, $M,v\not\models\neg H_i\bot\to p$ and, given $w\mathcal{K}_iv$, we also have $M,w\not\models K_i(\neg H_i\bot\to p)$. Thus, $M,w\not\models KH^{\to}$ for $\varphi=p$. □

Definition 3.3 *Class* $\mathcal{KH}$ of models for knowledge and hope consists of all Kripke models $M=\big((W,\mathcal{K}_1,\dots,\mathcal{K}_n,\mathcal{H}_1,\dots,\mathcal{H}_n),\pi\big)$ where (i) every $\mathcal{K}_i$ is an equivalence relation, (ii) every $\mathcal{H}_i$ is shift serial, and (iii) properties $\mathcal{H}$in$\mathcal{K}$ and one$\mathcal{H}$ are satisfied.

Proposition 3.4 *For all* $M=\big((W,\mathcal{K}_1,\dots,\mathcal{K}_n,\mathcal{H}_1,\dots,\mathcal{H}_n),\pi\big)\in\mathcal{KH}$, *each accessibility relation* $\mathcal{H}_i$ *is symmetric and transitive.*

Proof. To prove transitivity, let $w\mathcal{H}_iv$ and $v\mathcal{H}_iu$. Then $w\mathcal{K}_iv$ and $v\mathcal{K}_iu$ by $\mathcal{H}$in$\mathcal{K}$. Therefore, $w\mathcal{K}_iu$ by the transitivity of $\mathcal{K}_i$. $\mathcal{H}_i(w)\ni v$ is not empty. $\mathcal{H}_i(u)\neq\varnothing$ by the shift seriality of $\mathcal{H}_i$ because $v\mathcal{H}_iu$. Hence, $w\mathcal{H}_iu$ by one$\mathcal{H}$.

To prove symmetry, let $w\mathcal{H}_iv$. Then $w\mathcal{K}_iv$ by $\mathcal{H}$in$\mathcal{K}$. Therefore, $v\mathcal{K}_iw$ by the symmetry of $\mathcal{K}_i$. $\mathcal{H}_i(w)\ni v$ is not empty. $\mathcal{H}_i(v)\neq\varnothing$ by the shift seriality of $\mathcal{H}_i$ because $w\mathcal{H}_iv$. Hence, $v\mathcal{H}_iw$ by one$\mathcal{H}$. □

Remark 3.5 Hence, $\mathcal{H}_i$ are partial equivalence relations, so that property one$\mathcal{H}$ can be described as "no $\mathcal{K}_i$-equivalence class contains more than one $\mathcal{H}_i$-partial-equivalence class."

A natural way of obtaining the combined logic $\mathscr{KH}$ of hope and knowledge would be to combine the axioms and rules for hope from Fig. 2, standard S5 axioms and rules for knowledge, and connection axiom KH. Prop. 3.4, however, indicates that this would create redundancies. As we now show, in the presence of axiom KH, KB4 properties of hope originate from S5 properties of knowledge, albeit with the help of the translation of axiom $H=H_i\mathit{correct}_i$ from $\mathscr{H}_{\mathrm{co}}$ into language $\mathcal{L}_{KH}$. This translation $H^{\dagger}=H_i\neg H_i\bot$ can be called *necessary consistency* for hope and is known to be characterized by shift seriality. The resulting simplified axiom system is presented in Fig. 3.

Lemma 3.6 *For all $i \in \mathcal{A}$,*
(i) $\mathscr{KH} \vdash K_i\varphi \to H_i\varphi$;
(ii) $\mathscr{KH} \vdash H_i(\varphi \to \psi) \wedge H_i\varphi \to H_i\psi$;
(iii) *if* $\mathscr{KH} \vdash \varphi$, *then* $\mathscr{KH} \vdash H_i\varphi$.

Proof.

(i) $K_i\varphi \to K_i(\neg H_i\bot \to \varphi)$ is by the normality of K_i. From prop. tautology $K_i(\neg H_i\bot \to \varphi) \to \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big)$ and one direction $\big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big) \to H_i\varphi$ of KH, by syllogism, $K_i\varphi \to H_i\varphi$.

(ii)
1. $K_i\big(\neg H_i\bot \to (\varphi \to \psi)\big) \to \big(K_i(\neg H_i\bot \to \varphi) \to K_i(\neg H_i\bot \to \psi)\big)$ normality of K_i
2. $H_i(\varphi \to \psi) \to \Big(\neg H_i\bot \to K_i\big(\neg H_i\bot \to (\varphi \to \psi)\big)\Big)$ axiom KH
3. $H_i\varphi \to \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big)$ axiom KH
4. $H_i(\varphi \to \psi) \to \Big(H_i\varphi \to \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \psi)\big)\Big)$ from 1.–3.
5. $\big(\neg H_i\bot \to K_i(\neg H_i\bot \to \psi)\big) \to H_i\psi$ axiom KH
6. $H_i(\varphi \to \psi) \to (H_i\varphi \to H_i\psi)$ from 4.–5.

(iii) Easily follows from (i). □

Remark 3.7 Given that $K_i\varphi \to H_i\varphi$ is also known to characterize frame property $\mathcal{H}\text{in}\mathcal{K}$, one might ask whether $KH^{\leftarrow}$ is equivalent to $K_i\varphi \to H_i\varphi$. The answer is negative because $KH^{\leftarrow}$ only characterizes $\mathcal{H}\text{in}\mathcal{K}$ under the additional assumption of $\mathcal{H}_i$ being shift serial. For instance, consider a model M with $W = \{w, v\}$, $\mathcal{K}_j = W \times W$ for all $j \in \mathcal{A}$, $\mathcal{H}_j = W \times W$ for all $j \neq i$, non-shift-serial $\mathcal{H}_i = \{(w, v)\}$, $\pi(p) = \{w\}$ for some atom p, and $\pi(q) = W$ for all $q \in \mathsf{Prop} \setminus \{p\}$. Then $M, w \not\models \big(\neg H_i\bot \to K_i(\neg H_i\bot \to p)\big) \to H_ip$ but $M, w \models K_ip \to H_ip$.

Theorem 3.8 (Completeness of the logic of hope and knowledge) *$\mathscr{KH}$ is sound and complete with respect to $\mathcal{KH}$.*

Proof sketch. The soundness of KH follows because all axioms except for $KH^{\leftarrow}$ are frame characterizable and class $\mathcal{KH}$ consists of frames with corresponding properties, one of which is shift seriality, which takes care of the additional restriction for $KH^{\leftarrow}$. The normality of H_i is derived in Lemma 3.6. This enables us to use the standard canonical model $M^C = \big((W^C, \mathcal{K}_1^C, \ldots, \mathcal{K}_n^C, \mathcal{H}_1^C, \ldots, \mathcal{H}_n^C), \pi^C\big)$ construction for the logic and prove the Truth Lemma, i.e., that $M^C, \Gamma \models \varphi$ iff $\varphi \in \Gamma$ for each maximal $\mathscr{KH}$-consistent set $\Gamma \in W^C$. It remains to show that $M^C \in \mathcal{KH}$. The argument for $\mathcal{K}_i^C$ being equivalence relations and for $\mathcal{H}_i^C$ being shift serial is standard. Property $\mathcal{H}\text{in}\mathcal{K}$, i.e., $\mathcal{H}_i^C \subseteq \mathcal{K}_i^C$, easily follows from $\mathscr{KH} \vdash K_i\varphi \to H_i\varphi$ proved in Lemma 3.6. Thus, we only show property one$\mathcal{H}$, i.e., that $\Gamma\mathcal{K}_i^C\Delta$ implies $\Gamma\mathcal{H}_i^C\Delta$ whenever $\mathcal{H}_i^C(\Gamma) \neq \varnothing$ and $\mathcal{H}_i^C(\Delta) \neq \varnothing$ for any maximal consistent sets $\Gamma, \Delta \in W^C$. Assume $\Gamma\mathcal{K}_i^C\Delta$, $\mathcal{H}_i^C(\Gamma) \neq \varnothing$, and $\mathcal{H}_i^C(\Delta) \neq \varnothing$. Note that $\mathcal{H}_i^C(\Xi) \neq \varnothing$ implies $\neg H_i\bot \in \Xi$ due to the maximal consistency. We need to prove that $H_i\varphi \in \Gamma$ implies $\varphi \in \Delta$. If $H_i\varphi \in \Gamma$, then $\Gamma \vdash_{\mathscr{KH}} K_i(\neg H_i\bot \to \varphi)$ by axiom KH and

$\neg H_i\bot \in \Gamma$. Hence, $K_i(\neg H_i\bot \to \varphi) \in \Gamma$ and $\neg H_i\bot \to \varphi \in \Delta$ by the definition of $\mathcal{K}_i^C$. Since $\neg H_i\bot \in \Delta$, we conclude that $\varphi \in \Delta$ as required. □

Corollary 3.9 *$\mathscr{KH} \vdash H_i\varphi \to H_iH_i\varphi$ and $\mathscr{KH} \vdash \varphi \to H_i\neg H_i\neg\varphi$ for $i \in \mathcal{A}$.*

Proof. It immediately follows from Theorem 3.8 and Prop. 3.4. □

Definition 3.10 *Class $\mathcal{KH}^{n-f}$* consists of all Kripke models from $\mathcal{KH}$ that have all-but-f-serial frames with respect to $\mathcal{H}_i$ relations.

Corollary 3.11 *$\mathscr{KH} + Byz_f$ is sound and complete with respect to $\mathcal{KH}^{n-f}$.*

Logics $\mathscr{KH}$ and $\mathscr{KH} + Byz_f$ formalize both reliable (knowledge) and unreliable (hope) information agents possess in fault-tolerant distributed systems, the latter with at most f byzantine agents. The following two propositions outline the epistemic attitudes of agents who know that they are faulty and agents who know that they are correct.

Proposition 3.12 *$\mathscr{KH} \vdash K_iH_i\bot \to H_i\varphi$ for all $i \in \mathcal{A}$.*

Proof. We have $H_i\bot \to H_i\varphi$ is by the normality of H_i. From K_i-factivity $K_iH_i\bot \to H_i\bot$ we get $K_iH_i\bot \to H_i\varphi$ by syllogism. □

Proposition 3.13 *$\mathscr{KH} \vdash K_i\neg H_i\bot \to (H_i\varphi \leftrightarrow K_i\varphi)$ for all $i \in \mathcal{A}$.*

Proof. $\mathscr{KH} \vdash K_i\neg H_i\bot \to (K_i\varphi \to H_i\varphi)$ is an easy corollary of Prop. 3.6. We provide a derivation of $K_i\neg H_i\bot \to (H_i\varphi \to K_i\varphi)$. Firstly, we can obtain $K_i\neg H_i\bot \to \big(H_i\varphi \to K_i(\neg H_i\bot \to \varphi)\big)$ from K_i-factivity $K_i\neg H_i\bot \to \neg H_i\bot$ and direction $H_i\varphi \to \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big)$ of axiom KH by propositional reasoning. It remains to use K^K axiom $K_i(\neg H_i\bot \to \varphi) \to (K_i\neg H_i\bot \to K_i\varphi)$ to get $K_i\neg H_i\bot \to (H_i\varphi \to K_i\varphi)$ by propositional reasoning. □

We now turn to properties relevant for analyzing distributed systems. For instance, due to our earlier discussion of unsolvability of many distributed problems if too many agents become faulty and in view of the Knowledge of Preconditions Principle, it is typically necessary for agents to know that there are at least $n - f$ correct agents overall:

Proposition 3.14
- *$\mathscr{KH} + Byz_f \vdash K_iByz_f$ for all $i \in \mathcal{A}$.*
- *$\mathcal{KH}^{n-f} \models K_iByz_f$ for all $i \in \mathcal{A}$.*

Corollary 3.15 (In fault-free systems, hope is knowledge) *Recall that axiom Byz_0 rules out the presence of faulty agents. For any $i \in \mathcal{A}$,*

$$\mathscr{KH} + Byz_0 \vdash H_i\varphi \leftrightarrow K_i\varphi.$$

Proof. Follows from Remark 2.9 and Props. 3.13 and 3.14. □

Prop. 3.12 can be strengthened because a faulty agent hopes for anything even without knowing that it is faulty, i.e., $\mathscr{KH} \vdash H_i\bot \to H_i\varphi$. By contrast, for Prop. 3.13, the knowledge modality cannot be dropped: for a correct agent, hope does not yet mean knowledge, i.e., $\mathscr{KH} \nvdash \neg H_i\bot \to (H_i\varphi \leftrightarrow K_i\varphi)$. Instead, for a correct agent, hope turns out to be equivalent to another epistemic

attitude called *belief*, also used in distributed computing [22]. We introduce the following abbreviations for belief B_i, as well as for mutual knowledge E^K_G, mutual belief E^B_G, and mutual hope E^H_G among a group $G \subseteq \mathcal{A}$ of agents:

$$B_i\varphi := K_i(\neg H_i\bot \to \varphi) \qquad E^K_G\varphi := \bigwedge_{i\in G} K_i\varphi$$

$$E^B_G\varphi := \bigwedge_{i\in G} B_i\varphi \qquad E^H_G\varphi := \bigwedge_{i\in G} H_i\varphi$$

Note that this belief B_i is of type $\mathsf{K45}$ rather than $\mathsf{KD45}$.

Remark 3.16 It immediately follows that

$$\mathscr{KH} \vdash \neg H_i\bot \to (H_i\varphi \leftrightarrow B_i\varphi). \tag{5}$$

Thus, it might seem that, for correct agents, the use of hope is superficial and can be replaced by the better studied belief. The subtlety lies in the fact that belief is actionable in the sense of the Knowledge of Preconditions Principle, because the agent always knows its beliefs due to the positive introspection of knowledge. The same is not true regarding hope. And while hope of a correct agent i is equivalent to its belief, agents might not be aware of this equivalence if they are uncertain whether i is, in fact, correct.

We now show, by purely modal means, that in distributed systems with bounded number of faulty agents, mutual belief among a sufficiently large group of agents can be extracted from mutual hope among all agents. This ability to lift information received about hopes of all agents into actionable beliefs of a critical mass of correct agents is at the core of many distributed algorithms, including Firing Rebels [11] (albeit in a more complex temporal setting).

Proposition 3.17 $\mathscr{KH} + Byz_f \quad \vdash \quad E^H_{\mathcal{A}}\varphi \to \bigvee_{\substack{G\subseteq\mathcal{A}\\ |G|=n-f}} E^B_G\varphi.$

Proof. Follows from (5) and axiom Byz_f. □

Another indication of the independence of hope as an epistemic attitude is the fact that hope generally creates neither knowledge of hope nor hope of knowledge.

Proposition 3.18 (Knowledge and hope do not mix) *For any $i \in \mathcal{A}$,*

- *it is not the case that $\mathcal{KH} \models H_i\varphi \to H_iK_i\varphi$ for all $\varphi \in \mathcal{L}_{KH}$,*
- *it is not the case that $\mathcal{KH} \models H_i\varphi \to K_iH_i\varphi$ for all $\varphi \in \mathcal{L}_{KH}$.*

Proof. We use the same model to refute both statements but refute them for different formulas φ. Let $M = \big((W, \mathcal{K}_1, \dots, \mathcal{K}_n, \mathcal{H}_1, \dots, \mathcal{H}_n), \pi\big)$ such that $W = \{\mathsf{G}, \mathsf{B}\}$, $\mathcal{K}_j = W \times W$ for all $j \in \mathcal{A}$, $\mathcal{H}_i = \{(\mathsf{G}, \mathsf{G})\}$, $\mathcal{H}_j = W \times W$ for all $j \neq i$, and π be arbitrary. Now, all $\mathcal{K}_j$ are equivalence relations, all $\mathcal{H}_j$ are shift serial, $\mathcal{H}_j \subseteq \mathcal{K}_j$ for all $j \in \mathcal{A}$, and one$\mathcal{H}$ holds. In other words, $M \in \mathcal{KH}$. Clearly, agent i is correct in world G, i.e., $\mathcal{H}_i(\mathsf{G}) \neq \varnothing$, and faulty in world B, i.e., $\mathcal{H}_i(\mathsf{B}) = \varnothing$. We now have that, $M, \mathsf{G} \not\models H_i\neg H_i\bot \to H_iK_i\neg H_i\bot$ and $M, \mathsf{B} \not\models H_i\bot \to K_iH_i\bot$. □

Corollary 3.19 *For any $i \in \mathcal{A}$ and any $f > 0$,*
- *it is not the case that $\mathcal{KH}^{n-f} \models H_i\varphi \to H_iK_i\varphi$ for all $\varphi \in \mathcal{L}_{KH}$,*
- *it is not the case that $\mathcal{KH}^{n-f} \models H_i\varphi \to K_iH_i\varphi$ for all $\varphi \in \mathcal{L}_{KH}$.*

We finish this section by showing that designated atoms *correct*$_i$, which can be defined away via the hope modality as $\neg H_i\bot$, are not definable in language $\mathcal{L}_K := \mathcal{L}(\mathsf{Prop}, K_1, \ldots, K_n)$ with knowledge modalities only.

Definition 3.20 For a language $\mathcal{L} = \mathcal{L}(P, \heartsuit_1, \ldots, \heartsuit_m)$, let Kripke models $M = \big((W, R_1, \ldots, R_m), \pi\big)$ and $M' = \big((W', R'_1, \ldots, R'_m), \pi'\big)$ be given. A non-empty relation $Z \subseteq W \times W'$ is a *bisimulation* between M and M', notation $Z\colon M \leftrightarrows M'$, if for all wZw' and $j \in \{1, \ldots, m\}$:

atoms $w \in \pi(p)$ iff $w' \in \pi'(p)$ for all $p \in P$;

forth if wR_jv, then there is a $v' \in W'$ such that $w'R'_jv'$ and vZv';

back if $w'R'_jv'$, then there is a $v \in W$ such that wR_jv and vZv'.

We write $M \leftrightarrows M'$ if there is a bisimulation $Z\colon M \leftrightarrows M'$. We write $(M, w) \leftrightarrows (M', w')$ if there is a bisimulation $Z\colon M \leftrightarrows M'$ such that wZw'.

A *restricted (to Q) bisimulation* Z^Q is a bisimulation that satisfies **atoms** for all atoms $Q \subseteq P$, notation $Z^Q\colon M \leftrightarrows^Q M'$. And, similarly, $(M, w) \leftrightarrows^Q (M', w')$ means that wZ^Qw' for some such Z^Q.

Given $Q \subseteq P$, let us write $\mathcal{L}|Q := \mathcal{L}(Q, \heartsuit_1, \ldots, \heartsuit_m)$ for the language restricted to atoms from Q only. Further, let us write
- $(M, w) \equiv (M', w')$ to mean 'for all $\varphi \in \mathcal{L}$, $M, w \models \varphi$ iff $M', w' \models \varphi$' and
- $(M, w) \equiv^Q (M', w')$ to mean 'for all $\varphi \in \mathcal{L}|Q$, $M, w \models \varphi$ iff $M', w' \models \varphi$.'

Theorem 3.21 ([2])
- $(M, w) \leftrightarrows (M', w')$ *implies* $(M, w) \equiv (M', w')$.
- $(M, w) \leftrightarrows^Q (M', w')$ *implies* $(M, w) \equiv^Q (M', w')$.

In language $\mathcal{L}_{KH}$ with knowledge and hope, *correct*$_i$ is definable as $\neg H_i\bot$, making *faulty*$_i = \neg$*correct*$_i$ equivalent to $H_i\bot$. More precisely, every formula in language $\mathcal{L}(\mathsf{Prop} \sqcup \mathsf{Co}, K_1, ..., K_n, H_1, ..., H_n)$ is equivalent to a formula in language $\mathcal{L}_{KH}$ via the †-translation (cf. Lemma 2.7).

We now show that atoms *correct*$_i$ are not definable in the language with modalities for knowledge only.

Proposition 3.22 *Correctness of agents is not definable from knowledge.*

Proof. We now consider languages $\mathcal{L}_K^{\mathrm{co}} := \mathcal{L}(\mathsf{Prop} \sqcup \mathsf{Co}, K_1, ..., K_n)$ and $\mathcal{L}_K := \mathcal{L}_K^{\mathrm{co}}|\mathsf{Prop} = \mathcal{L}(\mathsf{Prop}, K_1, ..., K_n)$. Assume towards a contradiction that there is a $\varphi_i \in \mathcal{L}_K$ such that φ_i is equivalent to *correct*$_i$. Now consider (M, w) and (M', w') (for the set $\mathsf{Prop} \sqcup \mathsf{Co}$ of all propositions) such that $(M, w) \leftrightarrows^{\mathsf{Prop}} (M', w')$ but for some agent $i \in \mathcal{A}$ we have $w \in \pi($*correct*$_i)$ while, at the same time, $w' \notin \pi'($*correct*$_i)$. From Theorem 3.21 we obtain that $(M, w) \equiv^{\mathsf{Prop}} (M', w')$. Therefore, in particular $M, w \models \varphi_i$ iff $M', w' \models \varphi_i$. This contradicts that $M, w \models$ *correct*$_i$ but $M', w' \not\models$ *correct*$_i$. Thus, *correct*$_i$ is not definable in $\mathcal{L}_K$. □

Note that, were the models M and M' above to also have hope relations, then $w \in \pi(correct_i)$ and $w' \notin \pi'(correct_i)$ would imply that $\mathcal{H}_i(w) \neq \varnothing$ whereas $\mathcal{H}'_i(w') = \varnothing$, thus, precluding their bisimilarity restricted to Prop.

4 Modal Representation of Byzantine Behaviors, including Brain-in-a-Vat

The goal of this section is to show the utility of logic $\mathscr{KH}$ of hope and knowledge by providing axiomatic and semantic descriptions of several properties that originate from earlier analyses of fully byzantine distributed systems in [16]. Roughly speaking, an agent is called *fully byzantine* if neither its behavior nor its perceptions are restricted in any way. We already discussed axiom Byz_f restricting the number of such agents to at most f to ensure solvability of distributed problems. Another important feature of distributed systems with fully byzantine agents is that agents can never exclude the possibility of their perceptions being completely fabricated in a so-called *brain-in-a-vat scenario*. For details of the modeling via global runs in fault-tolerant distributed message-passing systems, which also include a temporal component, we refer the reader to [16]. In particular, the following properties were demonstrated:

- Agents cannot reliably establish their own correctness, as formalized by the axiom, for all $i \in \mathcal{A}$,

$$iByz \qquad := \qquad \neg K_i \neg H_i \bot.$$

- A faulty agent lacks any reliable information about other agents.[3] In particular, we generally assume that a faulty agent has no reliable information to decide whether any other agent is correct or faulty, as formalized by the axiom, for all $i \neq j$,

$$BiV \qquad := \qquad H_i \bot \to \neg K_i H_j \bot \wedge \neg K_i \neg H_j \bot.$$

From these two principles, by purely modal means, we can derive that no agent knows whether other agents are correct or faulty, as proved in [16] by complex manipulations of distributed system runs.

Proposition 4.1 *$\mathscr{KH} + iByz + BiV \vdash anyByz_{ij} \wedge anyCor_{ij}$ for all $i \neq j$ where*

$$anyByz_{ij}: \qquad \neg K_i \neg H_j \bot; \qquad\qquad anyCor_{ij}: \qquad \neg K_i H_j \bot.$$

Proof. $\neg K_i \neg\, H_i \bot \to \neg K_i \neg\, \neg K_i H_j \bot$ for any $i \neq j$ follows from BiV by propositional reasoning and the normality of K_i (applied to its dual $\neg K_i \neg$). Hence, $\neg K_i \neg H_i \bot \to \neg K_i H_j \bot$ by T^K. Thus, by invoking $iByz$ and using MP we get $\neg K_i H_j \bot = anyCor_{ij}$. The argument for $anyByz_{ij}$ is analogous. □

Proposition 4.2 (Frame characterization of agents' fallibility) *Axiom $iByz$ for $i \in \mathcal{A}$ is characterized by the* i-may-aseriality *property*

[3] An important exception here must be made for *a priori* knowledge of the system, e.g., logical laws, physical laws, or global specifications of the distributed system.

of frames $F = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{H}_1, \ldots, \mathcal{H}_n)$ *requiring that each world have a* $\mathcal{K}_i$*-indistinguishable world with no* $\mathcal{H}_i$*-outgoing arrows:*

$$(\forall w \in W)\big(\exists w' \in \mathcal{K}_i(w)\big) \quad \mathcal{H}_i(w') = \varnothing.$$

Proof. Take an arbitrary frame $F = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{H}_1, \ldots, \mathcal{H}_n)$. We need to show that for $i \in \mathcal{A}$,

$$F \models iByz \qquad \Longleftrightarrow \qquad F \text{ is } i\text{-may-aserial.}$$

($\Longrightarrow$) We prove the contrapositive. If F is not i-may-aserial, there is some world $w \in W$ such that $\mathcal{H}_i(w') \neq \varnothing$ for all $w' \in \mathcal{K}_i(w)$. Independent of a valuation π, for $M = (F, \pi)$ we have $M, w' \models \neg H_i\bot$ for all $w' \in \mathcal{K}_i(w)$. Hence, we get $M, w \models K_i\neg H_i\bot$ and, hence, $F \not\models iByz$ for this i.
($\Longleftarrow$) Let F be i-may-aserial. Take an arbitrary $w \in W$. It now follows that there is $w' \in \mathcal{K}_i(w)$ such that $\mathcal{H}_i(w') = \varnothing$. Therefore, for $M = (F, \pi)$ with any π, we have $M, w' \models H_i\bot$ and $M, w \models \neg K_i\neg H_i\bot$. The validity of $iByz$ in F for this i follows because w and π were chosen arbitrarily. □

Proposition 4.3 (Frame characterization of brain-in-a-vat)
Axiom BiV *for* $i \neq j$ *is characterized by the* BiValence *property of frames* $F = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{H}_1, \ldots, \mathcal{H}_n)$

$$(\forall w \in W)\Big(\mathcal{H}_i(w) = \varnothing \Longrightarrow \big(\exists w', w'' \in \mathcal{K}_i(w)\big)\big(\mathcal{H}_j(w') \neq \varnothing \wedge \mathcal{H}_j(w'') = \varnothing\big)\Big).$$

Proof. Take any frame $F = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{H}_1, \ldots, \mathcal{H}_n)$. We need to show that for $i \neq j$

$$F \models BiV \qquad \Longleftrightarrow \qquad F \text{ is BiValent.}$$

($\Longrightarrow$) We prove the contrapositive. If F is not BiValent, there is some $w \in W$ such that $\mathcal{H}_i(w) = \varnothing$ but either $\mathcal{H}_j(w') = \varnothing$ for all $w' \in \mathcal{K}_i(w)$ or $\mathcal{H}_j(w'') \neq \varnothing$ for all $w'' \in \mathcal{K}_i(w)$. Independent of a valuation π, for $M = (F, \pi)$ we then have $M, w \models K_i\neg H_j\bot \vee K_iH_j\bot$ despite $M, w \models H_i\bot$, and, hence, $F \not\models BiV$ for these $i \neq j$.
($\Longleftarrow$) Let F be BiValent. Take an arbitrary $w \in W$ such that $\mathcal{H}_i(w) = \varnothing$. It now follows that there are $w' \in \mathcal{K}_i(w)$ such that $\mathcal{H}_j(w') \neq \varnothing$ and $w'' \in \mathcal{K}_i(w)$ such that $\mathcal{H}_j(w'') = \varnothing$. Therefore, for $M = (F, \pi)$ with any π, we can conclude $M, w \models \neg K_iH_j\bot \wedge \neg K_i\neg H_j\bot$ whenever $M, w \models H_i\bot$. The validity of BiV in F for these $i \neq j$ follows because w and π were chosen arbitrarily. □

We can also easily derive by purely modal means that brain-in-a-vat scenarios are not compatible with fault-free systems:

Proposition 4.4 $\mathscr{KH} + Byz_0 \vdash \neg iByz$ *for each* $i \in \mathcal{A}$.

Proof. By Prop. 3.14 and the normality of K_i, given that $Byz_0 \to \neg H_i\bot$ is a propositional tautology, we have $\mathscr{KH} + Byz_0 \vdash K_i\neg H_i\bot$ for each $i \in \mathcal{A}$. Deriving $\neg iByz$ is now a matter of propositional reasoning. □

Another interesting special case is $f = 1$ (with $n > 1$). On the one hand, half of BiV becomes derivable and, hence, redundant:

Proposition 4.5 *If any agent but no more than one can be faulty, then no agent can establish the faultiness of other agents: for all $i \neq j \in \mathcal{A}$,*

$$\mathscr{KH} + Byz_1 + iByz \quad \vdash \quad \neg K_i H_j \bot.$$

Proof. We have $H_i\bot \to \neg H_j\bot$ by Byz_1 for any $j \neq i$. Thus, we can conclude $\neg K_i\neg\, H_i\bot \to \neg K_i\neg\, \neg H_j\bot$ by the normality of K_i, i.e., $\neg K_i\neg H_i\bot \to \neg K_i H_j\bot$. Since $\neg K_i\neg H_i\bot$ is axiom $iByz$, we conclude $\neg K_i H_j\bot$ by MP. □

On the other hand, the other half of BiV leads to undesirable consequences:

Proposition 4.6 *For $f = 1$, the inability of a faulty agent to establish correctness of somebody else would lead to its inability to establish its own faultiness: for all $i \neq j \in \mathcal{A}$,*

$$\mathscr{KH} + Byz_1 + (H_i\bot \to \neg K_i\neg H_j\bot) \quad \vdash \quad \neg K_i H_i\bot.$$

Proof. A correct agent i considers its own correctness possible by T^K, i.e., $\neg H_i\bot \to \neg K_i H_i\bot$. Formula $H_i\bot \to \neg K_i\neg H_j\bot$ for at least one $j \neq i$ is an assumption. At the same time, $H_j\bot \to \neg H_i\bot$ by Byz_1. As before, $\neg K_i\neg H_j\bot \to \neg K_i H_i\bot$ follows by the normality of K_i, yielding implication $H_i\bot \to \neg K_i H_i\bot$ by syllogism. Since we have derived $\neg K_i H_i\bot$ from both assumptions $\neg H_i\bot$ and $H_i\bot$, we get $\neg K_i H_i\bot$ without any assumptions by propositional reasoning. □

Remark 4.7 Intuitively, if an agent establishes its own faultiness, which does not run afoul of $iByz$ and can be used, e.g., for self-repairing agents, then it will thereby establish the correctness of all other agents. It seems wrong to prohibit this by adopting the respective half of BiV, whereas the other half is derivable anyway. We, therefore, propose using $\mathscr{KH} + Byz_f + BiV + iByz$ for $f \geq 2$ but $\mathscr{KH} + Byz_1 + iByz$ for $f = 1$. (The case of $f = 0$, which can be axiomatized by $\mathscr{KH} + Byz_0$, is more efficiently dealt with in the standard epistemic language.)

5 Common Hope and Common Knowledge

In this section, we introduce the common hope modality by analogy with the common knowledge modality and explore their relationship. We start by extending language $\mathcal{L}_{KH}$ with unary modal operator C_G^H for common hope and unary modal operator C_G^K for common knowledge, where $\varnothing \neq G \subseteq \mathcal{A}$ is an arbitrary group of agents. We will denote this extended language by $\mathcal{L}_{KH}^C$. Similar to common knowledge among a group G, by common hope of φ we intuitively mean mutual hope that φ and mutual hope of mutual hope that φ, etc.:

$$C_G^H \qquad \leftrightsquigarrow \qquad E_G^H\varphi \wedge E_G^H E_G^H\varphi \wedge E_G^H E_G^H E_G^H\varphi \wedge \ldots$$

Axiom system $\mathscr{KHC}$ for common knowledge and common hope consists of all the axioms of $\mathscr{KH}$ (formulated for $\mathcal{L}_{KH}^C$ formulas) plus the following

axioms and inference rules for all $\varnothing \neq G \subseteq \mathcal{A}$ and all formulas $\varphi, \psi \in \mathcal{L}^C_{KH}$:

$$Mix^H: \quad C^H_G\varphi \to E^H_G(\varphi \wedge C^H_G\varphi); \qquad Mix^K: \quad C^K_G\varphi \to E^K_G(\varphi \wedge C^K_G\varphi);$$
$$Ind^H: \quad \frac{\psi \to E^H_G(\varphi \wedge \psi)}{\psi \to C^H_G\varphi}; \qquad Ind^K: \quad \frac{\psi \to E^K_G(\varphi \wedge \psi)}{\psi \to C^K_G\varphi}.$$

That common knowledge has the properties of individual knowledge is well-known. Still, it may be surprising that common hope has the properties of individual hope. (Recall that common belief does not have the properties of individual KD45 belief as it lacks negative introspection.) Proofs are standard and omitted.

Proposition 5.1 *For any formulas $\varphi, \psi \in \mathcal{L}^C_{KH}$ and any $\varnothing \neq G \subseteq \mathcal{A}$:*

$$\mathscr{KHC} \vdash C^H_G(\varphi \to \psi) \wedge C^H_G\varphi \to C^H_G\psi \qquad \mathscr{KHC} \vdash C^K_G(\varphi \to \psi) \wedge C^K_G\varphi \to C^K_G\psi$$
$$\mathscr{KHC} \vdash C^H_G\varphi \to C^H_GC^H_G\varphi \qquad \mathscr{KHC} \vdash C^K_G\varphi \to C^K_GC^K_G\varphi$$
$$\mathscr{KHC} \vdash \neg C^K_G\varphi \to C^K_G\neg C^K_G\varphi$$
$$\mathscr{KHC} \vdash \varphi \Longrightarrow \mathscr{KHC} \vdash C^H_G\varphi \qquad \mathscr{KHC} \vdash \varphi \Longrightarrow \mathscr{KHC} \vdash C^K_G\varphi$$
$$\mathscr{KHC} \vdash \varphi \to C^H_G\neg C^H_G\neg\varphi \qquad \mathscr{KHC} \vdash C^K_G\varphi \to \varphi$$

Proposition 5.2 $\mathscr{KHC} \vdash C^K_G\varphi \to C^H_G\varphi$.

Proof.
1. $C^K_G\varphi \to E^K_G(\varphi \wedge C^K_G\varphi)$ — axiom Mix^K
2. $E^K_G(\varphi \wedge C^K_G\varphi) \to E^H_G(\varphi \wedge C^K_G\varphi)$ — follows from Prop. 3.6
3. $C^K_G\varphi \to E^H_G(\varphi \wedge C^K_G\varphi)$ — by syllogism from 1. and 2.
4. $C^K_G\varphi \to C^H_G\varphi$ — by Ind^H from 3.

□

Formulas of $\mathcal{L}^C_{KH}$ are also evaluated on models from $\mathcal{KH}$, with the new clauses for the common knowledge and common hope added as follows:

Definition 5.3 For a model $\big((W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{H}_1, \ldots, \mathcal{H}_n), \pi\big) \in \mathcal{KH}$, we define

$$\mathcal{K}^C_G := \left(\bigcup_{i\in G} \mathcal{K}_i\right)^+, \qquad \mathcal{H}^C_G := \left(\bigcup_{i\in G} \mathcal{H}_i\right)^+,$$

where R^+ is the transitive (but not reflexive) closure of a relation R. Then we define $M, w \models C^K_G\varphi$ iff $M, v \models \varphi$ for all $v \in \mathcal{K}^C_G(w)$ and $M, w \models C^H_G\varphi$ iff $M, v \models \varphi$ for all $v \in \mathcal{H}^C_G(w)$.

Theorem 5.4 (Completeness for common hope and knowledge) *$\mathscr{KHC}$ is sound and complete with respect to $\mathcal{KH}$.*

Proof sketch. The proof uses a finite version of the canonical model construction with maximal consistent sets restricted to subsets of an appropriately chosen *finite* set $cl(\varphi)$ of "extended" subformulas of a given formula φ

(cf. Fischer–Ladner closure [8]). Apart from the choice of this closure set, the proof is rather standard, if lengthy and technical. One shows that finite canonical model M^C_φ belongs to class $\mathcal{KH}$. Then the Truth Lemma is established. Finally, if $\mathscr{KHC} \nvdash \varphi$, then $\{\neg\varphi\}$ is consistent and can be extended to a world in M^C_φ, where φ is false. The main difficulty in this proof is finding the closure set $cl(\varphi)$, hence, we provide it below.

The closure set $cl(\varphi)$ for the finite canonical model is defined in several stages. We use $\mathcal{L}^b_i \subset \mathcal{L}^C_{KH}$ to denote all formulas *not* of the form $\neg H_i\bot \to \psi$.

- $cl_0(\varphi)$ is the smallest set that (a) contains φ and $H_i\neg H_i\bot$ for all $i \in \mathcal{A}$, (b) is closed under subformulas, and, for all $\psi \in \mathcal{L}^C_{KH}$ and $\varnothing \neq G \subseteq \mathcal{A}$, (c) contains $E^H_G(\psi \wedge C^H_G\psi)$ whenever $C^H_G\psi \in cl_0(\varphi)$ and (d) contains $E^K_G(\psi \wedge C^K_G\psi)$ whenever $C^K_G\psi \in cl_0(\varphi)$.
- $cl_1(\varphi) := cl_0(\varphi) \cup \{\neg\psi \mid \psi \in cl_0(\varphi)\}$.
- $cl_2(\varphi) := cl_1(\varphi) \cup \{H_i(\neg H_i\bot \to \psi) \mid K_i(\neg H_i\bot \to \psi) \in cl_1(\varphi)\} \cup \{K_i(\neg H_i\bot \to \psi) \mid H_i(\neg H_i\bot \to \psi) \in cl_1(\varphi)\} \cup \{K_i(\neg H_i\bot \to \psi), H_i(\neg H_i\bot \to \psi), \neg H_i\bot \to \psi \mid H_i\psi \in cl_1(\varphi), \psi \in \mathcal{L}^b_i\} \cup \{H_i\psi, K_i(\neg H_i\bot \to \psi), H_i(\neg H_i\bot \to \psi), \neg H_i\bot \to \psi \mid K_i\psi \in cl_1(\varphi), \psi \in \mathcal{L}^b_i\}$
- $cl_3(\varphi) := cl_2(\varphi) \cup \{\neg\psi \mid \psi \in cl_2(\varphi)\}$.
- $cl_4(\varphi) := cl_3(\varphi) \cup \{K_iK_i\psi, H_iK_i\psi \mid K_i\psi \in cl_3(\varphi)\} \cup \{K_i\neg K_i\psi, H_i\neg K_i\psi \mid \neg K_i\psi \in cl_3(\varphi)\}$
- $cl(\varphi) := cl_4(\varphi) \cup \{\neg\psi \mid \psi \in cl_4(\varphi)\}$.

□

Corollary 5.5 (Decidability) *$\mathscr{KHC}$ is conservative over $\mathscr{KH}$. Both have the finite model property (FMP) and, hence, are decidable.*

Proof. If $\mathscr{KHC} \nvdash \varphi$, the finite canonical model M^C_φ from the proof of Theorem 5.4 serves as a finite countermodel. Thus, $\mathscr{KHC}$ has the FMP. If $\mathscr{KH} \nvdash \varphi$ for $\varphi \in \mathcal{L}_{KH}$, then $\mathcal{KH} \not\models \varphi$ by the completeness of $\mathscr{KH}$ and $\mathscr{KHC} \nvdash \varphi$ by the soundness of $\mathscr{KHC}$. This proves the conservativity, which implies the FMP for $\mathscr{KH}$. □

So far, by and large, the relationship between common knowledge and common hope exhibited the same traits as between their individual variants. But the naive generalization of connection axiom KH is invalid for the common modalities (for at least two agents). We recall that $KH^\to$ corresponds to property one$\mathcal{H}$ that each knowledge equivalence class contains at most one hope partial equivalence class (Prop. 4). It is easy to see that when lifted to the common modalities, each common knowledge equivalence class may contain more than one common hope (partial) equivalence class, thus, invalidating the generalization. The proof of the proposition below provides a simple four-world countermodel to that effect:

Proposition 5.6 $\mathcal{KH} \not\models C^H_G p \leftrightarrow \big(\neg C^H_G\bot \to C^K_G(\neg C^H_G\bot \to p)\big)$ *if* $|G| \geq 2$.

Proof. To show this, we construct a countermodel from $\mathcal{KH}$. Let $i \neq j \in G$. Consider a Kripke model $M = \big((W, \mathcal{K}_1, \ldots, \mathcal{K}_n, \mathcal{H}_1, \ldots, \mathcal{H}_n), \pi\big) \in \mathcal{KH}$ such that

- $W = \{\mathsf{G}', \mathsf{G}'', \mathsf{B}', \mathsf{B}''\}$;
- $\mathcal{K}_i$ splits W into equivalence classes $\{\mathsf{G}', \mathsf{B}'\}$ and $\{\mathsf{G}'', \mathsf{B}''\}$;
- $\mathcal{K}_j$ splits W into equivalence classes $\{\mathsf{G}', \mathsf{B}''\}$ and $\{\mathsf{G}'', \mathsf{B}'\}$;
- $\mathcal{K}_l = \mathcal{K}_i$ for all $l \in G \setminus \{i, j\}$;
- all agents from G are faulty in bad worlds B' and B'' and correct in good worlds G' and G'', i.e., given conditions $\mathcal{H}\text{in}\mathcal{K}$ and one$\mathcal{H}$, partial equivalence relations $\mathcal{H}_l = \{(\mathsf{G}', \mathsf{G}'), (\mathsf{G}'', \mathsf{G}'')\}$ for all $l \in G$;
- $\pi(p) = \{\mathsf{G}'\}$;
- other elements are arbitrary.

We now have the following: on the one hand, $M, \mathsf{G}' \models C_G^H p$ because, for any $l \in G$, the only world $\mathcal{H}_l$-accessible from G' is G' itself.

On the other hand, $M, w \models C_G^H \bot$ iff $w \in \{\mathsf{B}', \mathsf{B}''\}$. In particular, we have $M, \mathsf{G}' \models \neg C_G^H \bot$ and $M, \mathsf{G}'' \models \neg C_G^H \bot$. Thus, $M, \mathsf{G}'' \not\models \neg C_G^H \bot \to p$ and, consequently, $M, \mathsf{G}' \not\models C_G^K(\neg C_G^H \bot \to p)$. Overall, we can conclude that $M, \mathsf{G}' \not\models \neg C_G^H \bot \to C_G^K(\neg C_G^H \bot \to p)$.

Thus, $\mathcal{KH} \not\models C_G^H p \leftrightarrow \big(\neg C_G^H \bot \to C_G^K(\neg C_G^H \bot \to p)\big)$. □

6 Related Works

It is interesting to observe that many of the "usual suspects" for an epistemic logic do not fit the properties of hope observed in the runs and systems modeling of Firing Rebels with Relay [11] and derived from the properties of knowledge in this paper. For instance, all extensions of $\mathsf{S4}_n$ are ruled out because hope (of faulty agents) is not factive, i.e., $\not\models H_i\varphi \to \varphi$. Similarly, $\mathsf{KD45}_n$ cannot be used because we take the inconsistency statement $H_i\bot$ to be the definition of agent i's faultiness rather than summarily ruling it out by axiom D.

It is notable that, independently, based on algebraic topological modeling, Goubault et al. [13] proposed $\mathsf{KB4}_n$ as an epistemic attitude for synchronous systems where agent malfunctions are restricted to *crash failures*. They call their $\mathsf{KB4}$ modalities 'knowledge,' use K_i for them, and define a dead agent as $K_i\bot$. We call our $\mathsf{KB4}$ modalities 'hope,' use H_i for them, and define an incorrect agent as $H_i\bot$, whereas 'knowledge' for us is a separate modality of type $\mathsf{S5}$. Our derivation of $\mathsf{KB4}$ properties of hope from the standard $\mathsf{S5}$ properties of knowledge helps explain the similarities between their findings for synchronous agents with at most crash failures and the system for fully byzantine asynchronous agents from [10]. This suggests $\mathsf{KB4}_n$ to be a good epistemic basis for studying a wide range of fault-tolerant systems.

Moses and Shoham [22] introduce three binary modal operators describing a single agent's beliefs as a form of knowledge relativized to an assumption (without committing to any type of knowledge or to any particular assumption). The most relevant of the three for us is the first one $B_1^\alpha\varphi := K(\alpha \to \varphi)$, where α is any formula.[4] Thus, dropping the agent subscript for a single agent, our notion of belief $B\varphi = K(\neg H\bot \to \varphi)$, see also [11, 15–17], coincides with their $B_1^{\neg H\bot}\varphi$.

[4] Here subscript 1 means "first operator out of three" rather than agent 1.

Bolander et al. [3] consider a version of public announcement logic (PAL), called *attention-based announcement logic*, where agents need not pay attention to a public announcement. Not being attentive could be viewed as a special type of fault, which is modeled in [3] by designated atoms h_i for each agent i. Thus, much like the knowledge of our byzantine agents depends on whether they are correct, i.e., whether $\neg H_i\bot$ is true, the knowledge of their agents after a public announcement depends on whether h_i is true. Another common concern is agents' introspective properties regarding their faults, with [3] considering systems for both non-fault-introspective and fault-introspective agents, the latter stipulating the *attention introspection property*: an attentive agent believes to be attentive, $h_i \to B_i h_i$, and an inattentive agent believes to be inattentive, $\neg h_i \to B_i \neg h_i$. This results in logic K_n for non-fault-introspective agents and an extension of logic $\mathsf{K45}_n$ for fault-introspective ones. The distinction between [3] and our byzantine agents is their lack of axiom B corresponding to frame symmetry. Note that, by the very nature of their work, [3] deals with dynamic epistemic notions. The authors also introduce an adaptation of relativized common belief [1] called *attention-based relativized common belief* defined as the fixpoint of the equation $x = E^{\chi}_{\mathcal{A}}(\varphi \wedge x)$, where $E^{\chi}_{\mathcal{A}} := \bigwedge_{i\in\mathcal{A}}\big(h_i \to B_i(\chi \to \varphi)\big)$ and where χ is the relativization formula. This closely resembles our notion of mutual hope $E^{H}_{\mathcal{A}} = \bigwedge_{i\in\mathcal{A}}(\neg H_i\bot \to K_i\big(\neg H_i\bot \to \varphi)\big)$.

7 Conclusion and Future Work

We provided a description of epistemic views of agents in fault-tolerant distributed systems with fully byzantine agents by means of a multimodal logic with two types of modalities — hope and knowledge — and showed how system specifications and properties of such agents can be represented by frame-characterizable properties. This analysis yielded new insights, for instance, into the distinctions between the case of fault-tolerant systems with at most one vs. several byzantine agents. This distinction was already observed in [16] but the newly provided axiomatic representation explains which of the general properties of byzantine agents are violated when all but one agents are correct.

The extension of our completeness result to the case of common hope and common knowledge is paving the way for the complete analysis of the Firing Rebels with Relay problem, which involves a temporal dimension and relies on a temporal generalization of mutual hope, called *eventual mutual hope* [11]. As for the case of inattentive agents in [3], we also plan to introduce a dynamic component to our logic of byzantine agents, in order to formalize how communication in distributed systems affects agents' epistemic state depending on agents' correctness. We would also like to describe common hope as relativized common knowledge along the lines of [3], but the difficulty is that in our case formula χ would have to depend on i.

All these developments and extensions will be guided by the need to represent specific types of faults commonly considered in distributed systems.

Acknowledgments. We are grateful to the anonymous reviewers for their valuable suggestions on how to improve the paper. Following one suggestion,

we have included decidability results. We are also grateful for multiple fruitful discussions with and/or enthusiastic support of Giorgio Cignarale, Rojo Randrianomentsoa, Hugo Rincón Galeana, Thomas Schlögl, and Ulrich Schmid.

References

[1] van Benthem, J., J. van Eijck and B. Kooi, *Logics of communication and change*, Information and Computation **204** (2006), pp. 1620–1662.
URL https://doi.org/10.1016/j.ic.2006.04.006

[2] Blackburn, P., M. de Rijke and Y. Venema, "Modal Logic," Cambridge Tracts in Theoretical Computer Science **53**, Cambridge University Press, 2001.
URL https://doi.org/10.1017/CBO9781107050884

[3] Bolander, T., H. van Ditmarsch, A. Herzig, E. Lorini, P. Pardo and F. Schwarzentruber, *Announcements to attentive agents*, Journal of Logic, Language and Information **25** (2016), pp. 1–35.
URL http://dx.doi.org/10.1007/s10849-015-9234-3

[4] Chandra, T. D., V. Hadzilacos and S. Toueg, *The weakest failure detector for solving consensus*, Journal of the ACM **43** (1996), pp. 685–722.
URL https://doi.org/10.1145/234533.234549

[5] Dolev, D., M. Függer, M. Posch, U. Schmid, A. Steininger and C. Lenzen, *Rigorously modeling self-stabilizing fault-tolerant circuits: An ultra-robust clocking scheme for systems-on-chip*, Journal of Computer and System Sciences **80** (2014), pp. 860–900.
URL https://doi.org/10.1016/j.jcss.2014.01.001

[6] Dolev, D., J. Y. Halpern and H. R. Strong, *On the possibility and impossibility of achieving clock synchronization*, Journal of Computer and System Sciences **32** (1986), pp. 230–250.
URL https://doi.org/10.1016/0022-0000(86)90028-0

[7] Fagin, R., J. Y. Halpern, Y. Moses and M. Y. Vardi, "Reasoning About Knowledge," MIT Press, 1995.

[8] Fischer, M. J. and R. E. Ladner, *Propositional dynamic logic of regular programs*, Journal of Computer and System Sciences **18** (1979), pp. 194–211.
URL https://doi.org/10.1016/0022-0000(79)90046-1

[9] Fischer, M. J., N. A. Lynch and M. Merritt, *Easy impossibility proofs for distributed consensus problems*, Distributed Computing **1** (1986), pp. 26–39.
URL https://doi.org/10.1007/BF01843568

[10] Fruzsa, K., *Hope for epistemic reasoning with faulty agents!*, in: *ESSLLI 2019 Student Session* (2019).
URL http://esslli2019.folli.info/wp-content/uploads/2019/08/tentative_proceedings.pdf

[11] Fruzsa, K., R. Kuznets and U. Schmid, *Fire!*, in: J. Halpern and A. Perea, editors, *Proceedings Eighteenth Conference on Theoretical Aspects of Rationality and Knowledge*, Electronic Proceedings in Theoretical Computer Science **335** (2021), pp. 139–153.
URL http://dx.doi.org/10.4204/EPTCS.335.13

[12] Függer, M. and U. Schmid, *Reconciling fault-tolerant distributed computing and systems-on-chip*, Distributed Computing **24** (2012), pp. 323–355.
URL http://dx.doi.org/10.1007/s00446-011-0151-7

[13] Goubault, É., J. Ledent and S. Rajsbaum, *A simplicial model for* $\mathbf{KB4_n}$*: Epistemic logic with agents that may die*, in: P. Berenbrink and B. Monmege, editors, *39th International Symposium on Theoretical Aspects of Computer Science (STACS 2022)*, Leibniz International Proceedings in Informatics (LIPIcs) **219** (2022), pp. 33:1–33:20.
URL https://doi.org/10.4230/LIPIcs.STACS.2022.33

[14] Halpern, J. Y. and Y. Moses, *Knowledge and common knowledge in a distributed environment*, Journal of the ACM **37** (1990), pp. 549–587.
URL https://doi.org/10.1145/79147.79161

[15] Kuznets, R., L. Prosperi, U. Schmid and K. Fruzsa, *Causality and epistemic reasoning in byzantine multi-agent systems*, in: L. S. Moss, editor, *Proceedings Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge*, Electronic Proceedings in Theoretical Computer Science **297** (2019), pp. 293–312.
URL http://dx.doi.org/10.4204/EPTCS.297.19

[16] Kuznets, R., L. Prosperi, U. Schmid and K. Fruzsa, *Epistemic reasoning with byzantine-faulty agents*, in: A. Herzig and A. Popescu, editors, *Frontiers of Combining Systems, 12th International Symposium, FroCoS 2019, London, UK, September 4–6, 2019, Proceedings*, Lecture Notes in Artificial Intelligence **11715** (2019), pp. 259–276.
URL http://dx.doi.org/10.1007/978-3-030-29007-8_15

[17] Kuznets, R., L. Prosperi, U. Schmid, K. Fruzsa and L. Gréaux, *Knowledge in Byzantine message-passing systems I: Framework and the causal cone*, Technical Report TUW-260549, TU Wien (2019).
URL https://publik.tuwien.ac.at/files/publik_260549.pdf

[18] Lamport, L., R. Shostak and M. Pease, *The Byzantine Generals Problem*, ACM Transactions on Programming Languages and Systems **4** (1982), pp. 382–401.
URL http://dx.doi.org/10.1145/357172.357176

[19] Mendes, H., C. Tasson and M. Herlihy, *Distributed computability in Byzantine asynchronous systems*, in: *STOC 2014: 46th Annual Symposium on the Theory of Computing* (2014), pp. 704–713.
URL http://dx.doi.org/10.1145/2591796.2591853

[20] Mitchell, J. C. and E. Moggi, *Kripke-style models for types lambda calculus*, Annals of Pure and Applied Logic **51** (1991), pp. 99–124.
URL http://dx.doi.org/10.1016/0168-0072(91)90067-V

[21] Moses, Y., *Relating knowledge and coordinated action: The Knowledge of Preconditions principle*, in: R. Ramanujam, editor, *Proceedings Fifteenth Conference on Theoretical Aspects of Rationality and Knowledge*, Electronic Proceedings in Theoretical Computer Science **215** (2016), pp. 231–245.
URL http://dx.doi.org/10.4204/EPTCS.215.17

[22] Moses, Y. and Y. Shoham, *Belief as defeasible knowledge*, Artificial Intelligence **64** (1993), pp. 299–321.
URL https://doi.org/10.1016/0004-3702(93)90107-M

[23] Pessin, A. and S. Goldberg, editors, "The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's the "Meaning of 'Meaning'"," Routledge, 1996.
URL https://doi.org/10.4324/9781315284811

[24] Robinson, P. and U. Schmid, *The Asynchronous Bounded-Cycle model*, Theoretical Computer Science **412** (2011), pp. 5580–5601.
URL https://doi.org/10.1016/j.tcs.2010.08.001

[25] Schlögl, T., U. Schmid and R. Kuznets, *The persistence of false memory: Brain in a vat despite perfect clocks*, in: T. Uchiya, Q. Bai and I. Marsá Maestre, editors, *PRIMA 2020: Principles and Practice of Multi-Agent Systems: 23rd International Conference, Nagoya, Japan, November 18–20, 2020, Proceedings*, Lecture Notes in Artificial Intelligence **12568**, Springer, 2021 pp. 403–411.
URL https://doi.org/10.1007/978-3-030-69322-0_30

[26] Srikanth, T. K. and S. Toueg, *Optimal clock synchronization*, Journal of the ACM **34** (1987), pp. 626–645.
URL http://dx.doi.org/10.1145/28869.28876

[27] Srikanth, T. K. and S. Toueg, *Simulating authenticated broadcasts to derive simple fault-tolerant algorithms*, Distributed Computing **2** (1987), pp. 80–94.
URL http://dx.doi.org/10.1007/BF01667080

[28] Widder, J. and U. Schmid, *The Theta-Model: achieving synchrony without clocks*, Distributed Computing **22** (2009), pp. 29–47.
URL http://dx.doi.org/10.1007/s00446-009-0080-x