Framing Faultiness Kripke Style

Roman Kuznets
TU Wien

joint work with Hans van Ditmarsch and Krisztina Fruzsa
Full paper published as "A New Hope," AiML 2022

LATD 2022 AND MOSAIC KO
September 5–10, 2022
Paestum, Italy

Plan of the talk

1. New epistemic modality hope
2. New axiom system for hope
3. Frame conditions for properties of distributed systems

"A New Hope"

What is hope?

Hope is an epistemic[a] modality for analyzing fault-tolerant distributed systems.

[a] epistemic/doxastic

"A New Hope"

What is hope?

Hope is an epistemic[a] modality for analyzing fault-tolerant distributed systems.

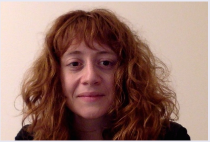
[a]epistemic/doxastic

Why is hope?

- belief — what agents think
- knowledge — belief when agents are right
- hope — ???

Hope, what is it good for? Faultless edition

Mental experiment #1



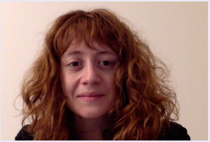
What do I learn when I read Sonia Marin's completeness proof for ecumenical modal logic EML?

- Does Sonia know that EML is complete?
- Do I know that Sonia knows that EML is complete?
- Do I know that EML is complete?

Hope, what is it good for? Faultless edition

Mental experiment #1

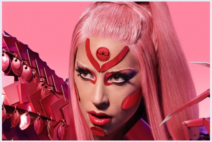
What do I learn when I read Sonia Marin's completeness proof for ecumenical modal logic EML?



- Does Sonia know that EML is complete? $K_s Co$
- Do I know that Sonia knows that EML is complete? $K_i K_s Co$
- Do I know that EML is complete? $K_i Co$

Hope, what is it good for? Fault-tolerant edition

Mental experiment #2

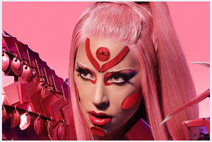


What do I learn when I read
Lady Gaga's proof that $P \neq NP$?

- Does Lady Gaga know $P \neq NP$?
- Do I know that Lady Gaga knows $P \neq NP$?
- Do I know $P \neq NP$?

Hope, what is it good for? Fault-tolerant edition

Mental experiment #2

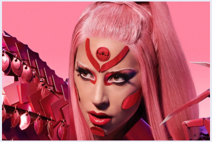


What do I learn when I read
Lady Gaga's proof that $P \neq NP$?

- Does Lady Gaga know $P \neq NP$? — probably not
- Do I know that Lady Gaga knows $P \neq NP$? — $\neg K_i K_{lg} Ne$
- Do I know $P \neq NP$? — $\neg K_i Ne$

Hope, what is it good for? Fault-tolerant edition

Mental experiment #2



What do I believe after I read
Lady Gaga's proof that $P \neq NP$?

- Does Lady Gaga believe $P \neq NP$? maybe?
- Do I believe that Lady Gaga believes $P \neq NP$? $\neg B_i B_{lg} Ne$
- Do I believe $P \neq NP$? no thanks to Lady Gaga

Is communication in fault-tolerant systems useless?

Knowledge of Preconditions Principle, KoP (Moses, 2015)

If φ is a necessary condition for agent i performing an action, then $K_i\varphi$ is also a necessary condition for this action.

Is communication in fault-tolerant systems useless?

Knowledge of Preconditions Principle, KoP (Moses, 2015)

If φ is a necessary condition for agent i performing an action, then $K_i\varphi$ is also a necessary condition for this action.

Corollary

If communication does not change the epistemic state of i, it cannot affect i's actions.

Belief as Knowledge Relative to Correctness

Belief as defeasible knowledge (Moses and Shoham, 1993)

$$B_i\varphi \quad := \quad K_i(correct_i \to \varphi)$$

The only non-factive situations are when i is faulty.

Belief as Knowledge Relative to Correctness

Belief as defeasible knowledge (Moses and Shoham, 1993)

$$B_i\varphi \quad := \quad K_i(correct_i \to \varphi)$$

The only non-factive situations are when i is faulty.

Malfunctioning agents tell no lies

Suppose faulty agents may be mistaken but cannot lie.
Then agent i receiving message φ from agent j results in

$$B_i B_j \varphi$$

Belief as Knowledge Relative to Correctness

Belief as defeasible knowledge (Moses and Shoham, 1993)

$$B_i\varphi \quad := \quad K_i(correct_i \rightarrow \varphi)$$

The only non-factive situations are when i is faulty.

Malfunctioning agents tell no lies

Suppose faulty agents may be mistaken but cannot lie.

Then agent i receiving message φ from agent j results in

$$B_i B_j \varphi$$

Fully byzantine agents can lie maliciously

Belief is not sufficient: no reason to conclude $B_i B_j \varphi$.

Is there any hope to analyze fully byzantine agents?

Is there any hope to analyze fully byzantine agents?

Belief as defeasible knowledge (Moses and Shoham, 1993)

$$B_i\varphi \quad := \quad K_i(correct_i \rightarrow \varphi)$$

Our first hope (K, Prosperi, Schmid, and Fruzsa, 2019)

$$H_i\varphi \quad := \quad correct_i \rightarrow K_i(correct_i \rightarrow \varphi)$$

Is there any hope to analyze fully byzantine agents?

Belief as defeasible knowledge (Moses and Shoham, 1993)

$$B_i\varphi \quad := \quad K_i(correct_i \to \varphi)$$

Our first hope (K, Prosperi, Schmid, and Fruzsa, 2019)

$$H_i\varphi \quad := \quad correct_i \to K_i(correct_i \to \varphi)$$

Mental experiment #2 revisited

What do I learn when I read Lady Gaga's proof that $P \neq NP$?

$$B_i H_{lg} Ne$$

or

$$K_i\Big(correct_i \to \big(correct_{lg} \to K_{lg}(correct_{lg} \to Ne)\big)\Big)$$

Is there any hope to analyze fully byzantine agents?

Belief as defeasible knowledge (Moses and Shoham, 1993)

$$B_i\varphi \quad := \quad K_i(correct_i \to \varphi)$$

Our first hope (K, Prosperi, Schmid, and Fruzsa, 2019)

$$H_i\varphi \quad := \quad correct_i \to K_i(correct_i \to \varphi)$$

Mental experiment #2 revisited

What do I learn when I read Lady Gaga's proof that $P \neq NP$?

$$B_i H_{lg} Ne$$

or

$$K_i\Big(correct_i \to \big(correct_{lg} \to K_{lg}(correct_{lg} \to Ne)\big)\Big)$$

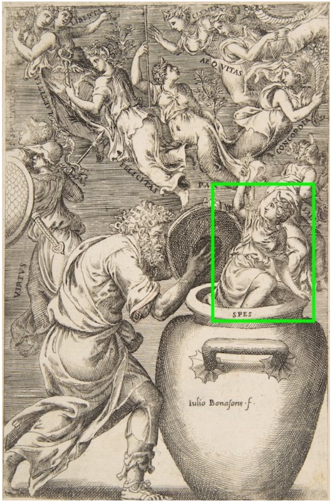
The outer knowledge operator K_i makes it a suitable necessary condition under KoP.

First Glimmers of Hope

We first identified hope modality

while analyzing a simplified version of the consistent broadcasting primitive, which is used for

- byzantine fault-tolerant clock synchronization,
- synchronous consensus,
- reduction of byzantine systems to systems with crash failures only.



Giulio Bonasone, *Epimetheus opening Pandora's box*

Fault-tolerant Distributed Systems with Fully Byzantine Agents

Message-passing distributed systems

- No central controller.
- Each agent has perfect recall but only local information.
- Information from other agents is exclusively via messages.

Fault-tolerant Distributed Systems with Fully Byzantine Agents

Message-passing distributed systems

- No central controller.
- Each agent has perfect recall but only local information.
- Information from other agents is exclusively via messages.

Messages can be

- lost
- delayed
- fake

in fault tolerant systems

Fault-tolerant Distributed Systems with Fully Byzantine Agents

Message-passing distributed systems

- No central controller.
- Each agent has perfect recall but only local information.
- Information from other agents is exclusively via messages.

Messages can be

- lost
- delayed
- fake — in fault tolerant systems

Fully byzantine agents can

- deviate from their protocol
- collude with each other in order to thwart the correct ones
- have false memories

Why We Have Hope: Executive summary

Hope is...

Why We Have Hope: Executive summary

Hope is...

- technically convenient

Why We Have Hope: Executive summary

Hope is...

- technically convenient
- weak enough to represent unreliable communication

Why We Have Hope: Executive summary

Hope is...

- technically convenient
- weak enough to represent unreliable communication
- enables to formulate system specification uniformly for correct and faulty agents:

whenever agent i acts, it must be that $H_i \varphi$

Our first hope, axiomatized

The language contains special propositional atoms $correct_i$:

$$\varphi ::= \bot \mid p \mid correct_i \mid (\varphi \to \varphi) \mid H_i\varphi$$

$$faulty_i \quad := \quad \neg correct_i \quad = \quad correct_i \to \bot$$

Our first hope, axiomatized

The language contains special propositional atoms $correct_i$:

$$\varphi ::= \bot \mid p \mid correct_i \mid (\varphi \to \varphi) \mid H_i\varphi$$

$$faulty_i \quad := \quad \neg correct_i \quad = \quad correct_i \to \bot$$

Axiomatic system $\mathscr{H}_{\mathrm{co}}$ (Fruzsa, 2019)

P : all propositional tautologies

K^H: $H_i(\varphi \to \psi) \to (H_i\varphi \to H_i\psi)$

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

5^H : $\neg H_i\varphi \to H_i\neg H_i\varphi$

MP: $\dfrac{\varphi \quad \varphi \to \psi}{\psi}$

T'^H : $correct_i \to (H_i\varphi \to \varphi)$

F : $faulty_i \to H_i\varphi$

H : $H_i correct_i$

Nec^H: $\dfrac{\varphi}{H_i\varphi}$

i.e., $\mathscr{H}_{\mathrm{co}} = \mathscr{K}45_n + T'^H + F + H$

Our first hope, axiomatized

The language contains special propositional atoms $correct_i$:

$$\varphi ::= \bot \mid p \mid correct_i \mid (\varphi \to \varphi) \mid H_i\varphi$$

$$faulty_i \quad := \quad \neg correct_i \quad = \quad correct_i \to \bot$$

Axiomatic system $\mathscr{H}_{\mathrm{co}}$ (Fruzsa, 2019)

P : all propositional tautologies

K^H: $H_i(\varphi \to \psi) \to (H_i\varphi \to H_i\psi)$

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

5^H : $\neg H_i\varphi \to H_i\neg H_i\varphi$

MP: $\dfrac{\varphi \quad \varphi \to \psi}{\psi}$

T'^H : $correct_i \to (H_i\varphi \to \varphi)$

F : $faulty_i \to H_i\varphi$

H : $H_i correct_i$

Nec^H: $\dfrac{\varphi}{H_i\varphi}$

i.e., $\mathscr{H}_{\mathrm{co}} = \mathscr{K}45_n + T'^H + F + H$

NB Not a normal modal logic.

Our first hope, Kripke style

Class $\mathcal{K}45_n^{\mathrm{co}}$: Kripke models with n transitive, euclidean relations $\mathcal{H}_1, \ldots, \mathcal{H}_n$. such that

1. $w \vDash correct_i \quad \Longrightarrow \quad w\mathcal{H}_i w,$
2. $w \nvDash correct_i \quad \Longrightarrow \quad \mathcal{H}_i(w) = \varnothing,$
3. $w\mathcal{H}_i w' \quad \Longrightarrow \quad w' \vDash correct_i.$

where $\mathcal{H}_i(w) := \{v \mid w\mathcal{H}_i v\}$.

Completeness Theorem (Fruzsa, 2019)

$\mathscr{H}_{\mathrm{co}}$ is sound and complete w.r.t. $\mathcal{K}45_n^{\mathrm{co}}$.

Our first hope, Kripke style

Class $\mathcal{K}45^{\mathrm{co}}_n$: Kripke models with n transitive, euclidean relations $\mathcal{H}_1, \dots, \mathcal{H}_n$. such that

1. $w \vDash correct_i \quad \Longrightarrow \quad w\mathcal{H}_i w$,
2. $w \nvDash correct_i \quad \Longrightarrow \quad \mathcal{H}_i(w) = \varnothing$,
3. $w\mathcal{H}_i w' \quad \Longrightarrow \quad w' \vDash correct_i$.

where $\mathcal{H}_i(w) := \{v \mid w\mathcal{H}_i v\}$.

Completeness Theorem (Fruzsa, 2019)

$\mathscr{H}_{\mathrm{co}}$ is sound and complete w.r.t. $\mathcal{K}45^{\mathrm{co}}_n$.

Downsides

- not normal
- no frame characterization
- redundant in presence of knowledge:
  $H_i\varphi = correct_i \to K_i(correct_i \to \varphi)$.

The moment of ~~Eureka~~ Hope

It happened one day in Heerlen

- $w \models correct_i \quad \Longrightarrow \quad w\mathcal{H}_i w \quad \Longrightarrow \quad \mathcal{H}_i(w) \neq \varnothing,$
- $w \not\models correct_i \quad \Longrightarrow \quad \mathcal{H}_i(w) = \varnothing,$

The moment of ~~Eureka~~ Hope

It happened one day in Heerlen

- $w \vDash correct_i \quad \Longrightarrow \quad w\mathcal{H}_i w \quad \Longrightarrow \quad \mathcal{H}_i(w) \neq \varnothing$,
- $w \nvDash correct_i \quad \Longrightarrow \quad \mathcal{H}_i(w) = \varnothing$,

Krisztina and Hans: "Hey, Roman, did you know that

$$w \vDash correct_i \quad \Longleftrightarrow \quad \mathcal{H}_i(w) \neq \varnothing$$

$$correct_i \quad \longleftrightarrow \quad \neg H_i \bot$$

The moment of ~~Eureka~~ Hope

It happened one day in Heerlen

- $w \vDash correct_i \quad \Longrightarrow \quad w\mathcal{H}_i w \quad \Longrightarrow \quad \mathcal{H}_i(w) \neq \varnothing,$
- $w \nvDash correct_i \quad \Longrightarrow \quad \mathcal{H}_i(w) = \varnothing,$

Krisztina and Hans: “Hey, Roman, did you know that

$w \vDash correct_i \quad \Longleftrightarrow \quad \mathcal{H}_i(w) \neq \varnothing$

$correct_i \quad \longleftrightarrow \quad \neg H_i \bot$

Roman to himself...

@#&*$ OMG, I should have seen this...

The moment of ~~Eureka~~ Hope

It happened one day in Heerlen

- $w \vDash correct_i \quad \Longrightarrow \quad w\mathcal{H}_i w \quad \Longrightarrow \quad \mathcal{H}_i(w) \neq \varnothing,$
- $w \nvDash correct_i \quad \Longrightarrow \quad \mathcal{H}_i(w) = \varnothing,$

Krisztina and Hans: "Hey, Roman, did you know that

$w \vDash correct_i \quad \Longleftrightarrow \quad \mathcal{H}_i(w) \neq \varnothing$

$correct_i \quad \longleftrightarrow \quad \neg H_i \bot$

Roman to himself...

@#&*$ OMG, I should have seen this...

Roman: "Deer Esteemed Colleagues,

Sounds very interesting. Good work. Let us continue this.

The NEW hope from Heerlen

Now in the standard multimodal language:

$$\varphi ::= \bot \mid p \mid (\varphi \to \varphi) \mid H_i\varphi$$

$$correct_i := \neg H_i\bot, \qquad faulty_i := H_i\bot$$

The NEW hope from Heerlen

Now in the standard multimodal language:

$$\varphi ::= \bot \mid p \mid (\varphi \to \varphi) \mid H_i\varphi$$

$$\mathit{correct}_i := \neg H_i\bot, \qquad \mathit{faulty}_i := H_i\bot$$

Axiomatic system $\mathscr{H}$ (van Ditmarsch, Fruzsa, K, 2022)

P : all propositional tautologies

K^H: $H_i(\varphi \to \psi) \to (H_i\varphi \to H_i\psi)$

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

B^H: $\varphi \to H_i\neg H_i\neg\varphi$

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

The NEW hope from Heerlen

Now in the standard multimodal language:

$$\varphi ::= \bot \mid p \mid (\varphi \to \varphi) \mid H_i\varphi$$

$$correct_i := \neg H_i\bot, \qquad faulty_i := H_i\bot$$

Axiomatic system $\mathscr{H}$ (van Ditmarsch, Fruzsa, K, 2022)

P : all propositional tautologies

K^H: $H_i(\varphi \to \psi) \to (H_i\varphi \to H_i\psi)$

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

B^H: $\varphi \to H_i\neg H_i\neg\varphi$

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

i.e., $\mathscr{H} = \mathscr{KB}4_n$ and is

- a normal modal logic,
- complete w.r.t. class $\mathcal{KB}4_n$ of frames with n transitive, symmetric relations.

Same hope, new axioms

New $\mathscr{H}$ and old $\mathscr{H}_{\mathrm{co}}$ are equivalent in the following sense:

$$\mathscr{H} \vdash \varphi \quad \Longrightarrow \quad \mathscr{H}_{\mathrm{co}} \vdash \varphi$$
$$\mathscr{H}_{\mathrm{co}} \vdash \varphi \quad \Longrightarrow \quad \mathscr{H} \vdash \varphi^{\dagger}$$

where $\varphi^{\dagger}$ is obtained by replacing

- each $correct_i$ in φ with $\neg H_i \bot$ and

Proper Language for Fault-Tolerant Distributed Systems

What we need

- knowledge K_i as the basis of agents' actions via KoP
- hope H_i to describe information accumulation

Proper Language for Fault-Tolerant Distributed Systems

What we need

- knowledge K_i as the basis of agents' actions via KoP
- hope H_i to describe information accumulation

What we gain for free

- correctness atoms $correct_i := \neg H_i \bot$
- belief $B_i \varphi := K_i(correct_i \to \varphi)$

Axioms of Hope and Knowledge

The language with 2 modalities for each agent:

$$\varphi ::= \bot \mid p \mid (\varphi \to \varphi) \mid K_i\varphi \mid H_i\varphi$$

$$\mathit{correct}_i := \neg H_i\bot, \qquad \mathit{faulty}_i := H_i\bot$$

Axioms of Hope and Knowledge

The language with 2 modalities for each agent:

$$\varphi ::= \bot \mid p \mid (\varphi \to \varphi) \mid K_i\varphi \mid H_i\varphi$$

$$correct_i := \neg H_i\bot, \qquad faulty_i := H_i\bot$$

Axiomatic system $\mathscr{KH}$

P : all propositional tautologies

$H^\dagger$: $H_i\neg H_i\bot$

K^K : $K_i(\varphi \to \psi) \wedge K_i\varphi \to K_i\psi$

4^K : $K_i\varphi \to K_iK_i\varphi$

5^K : $\neg K_i\varphi \to K_i\neg K_i\varphi$

T^K : $K_i\varphi \to \varphi$

MP: $\dfrac{\varphi \quad \varphi \to \psi}{\psi}$

Nec^K: $\dfrac{\varphi}{K_i\varphi}$

KH : $H_i\varphi \leftrightarrow \big(\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi)\big)$

Axioms of Hope and Knowledge

The language with 2 modalities for each agent:

$$\varphi ::= \bot \mid p \mid (\varphi \to \varphi) \mid K_i\varphi \mid H_i\varphi$$

$$correct_i := \neg H_i\bot, \qquad faulty_i := H_i\bot$$

Axiomatic system $\mathscr{KH}$

P : all propositional tautologies

$H^\dagger$: $H_i\neg H_i\bot$

K^K : $K_i(\varphi \to \psi) \wedge K_i\varphi \to K_i\psi$

4^K : $K_i\varphi \to K_iK_i\varphi$

5^K : $\neg K_i\varphi \to K_i\neg K_i\varphi$

T^K : $K_i\varphi \to \varphi$

MP: $\dfrac{\varphi \quad \varphi \to \psi}{\psi}$ $\qquad Nec^K$: $\dfrac{\varphi}{K_i\varphi}$

KH : $H_i\varphi \leftrightarrow (\neg H_i\bot \to K_i(\neg H_i\bot \to \varphi))$

i.e., $\mathscr{KH} = \mathscr{S}5_n^K + H^\dagger + KH$

Semantics of Hope and Knowledge

Completeness Theorem (van Ditmarsch, Fruzsa, K, 2022)

$\mathscr{KH}$ is sound and complete w.r.t. class $\mathcal{KH}$ of models

- with n equivalence relations $\mathcal{K}_i$ for knowledge modalities,
- with n shift-serial relations $\mathcal{H}_i$ for hope modalities (shift serial means $w\mathcal{H}_i v \Longrightarrow v\mathcal{H}_i v$),
- such that $w\mathcal{H}_i v \Longrightarrow w\mathcal{K}_i v$
- such that $\mathcal{H}_i(w) \neq \varnothing \wedge \mathcal{H}_i(v) \neq \varnothing \wedge w\mathcal{K}_i v \Longrightarrow w\mathcal{H}_i v$

In the class $\mathcal{KH}$

- $\mathcal{H}_i$ are partial equivalence relations, i.e., transitive and symmetric;
- each $\mathcal{K}_i$ cluster contains at most one $\mathcal{H}_i$ cluster.

Semantics of Hope and Knowledge

Completeness Theorem (van Ditmarsch, Fruzsa, K, 2022)

$\mathscr{KH}$ is sound and complete w.r.t. class $\mathcal{KH}$ of models

- with n equivalence relations $\mathcal{K}_i$ for knowledge modalities,
- with n shift-serial relations $\mathcal{H}_i$ for hope modalities
  (shift serial means $w\mathcal{H}_i v \Longrightarrow v\mathcal{H}_i v$),
- such that $w\mathcal{H}_i v \Longrightarrow w\mathcal{K}_i v$
- such that $\mathcal{H}_i(w) \neq \varnothing \wedge \mathcal{H}_i(v) \neq \varnothing \wedge w\mathcal{K}_i v \Longrightarrow w\mathcal{H}_i v$

In the class $\mathcal{KH}$

- $\mathcal{H}_i$ are partial equivalence relations,
  i.e., transitive and symmetric;
- each $\mathcal{K}_i$ cluster contains at most one $\mathcal{H}_i$ cluster.

- normal logic with frame characterization
- can express both *correct*$_i$ and Moses–Shoham's belief B_i

Distributed Properties Kripke Style

Curb Your Byzantiness

Typical distributed specification:

The number of byzantine agents in a run cannot exceed f out of n.

Usually

- $n \geq 2f + 1$ or
- $n \geq 3f + 1$.

Distributed Properties Kripke Style

Curb Your Byzantiness

Typical distributed specification:

The number of byzantine agents in a run cannot exceed f out of n.

Usually

- $n \geq 2f + 1$ or
- $n \geq 3f + 1$.

Axiom representation

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

Frame characterization

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

Brain in a Vat

Brain-in-a-Vat Lemma (K, Prosperi, Schmid, and Fruzsa, 2019)

No matter what it observed, no agent (whether correct or faulty), can ever rule out the possibility of those observations being artificially manufactured and not real.

Brain in a Vat

Brain-in-a-Vat Lemma (K, Prosperi, Schmid, and Fruzsa, 2019)

No matter what it observed, no agent (whether correct or faulty), can ever rule out the possibility of those observations being artificially manufactured and not real.

If $f \geq 1$, i.e., if at least one agent can become byzantine, no agent can ever know that

- a particular action or event actually happened;
- it itself is correct;
- another agent is byzantine.

Brain in a Vat

Brain-in-a-Vat Lemma (K, Prosperi, Schmid, and Fruzsa, 2019)

No matter what it observed, no agent (whether correct or faulty), can ever rule out the possibility of those observations being artificially manufactured and not real.

If $f \geq 1$, i.e., if at least one agent can become byzantine, no agent can ever know that

- a particular action or event actually happened;
- it itself is correct;
- another agent is byzantine.

If $f \geq 2$, i.e., if more than one agent can become byzantine, no agent can ever know that

- another agent is correct.

Brain in a Vat

Brain-in-a-Vat Lemma (K, Prosperi, Schmid, and Fruzsa, 2019)

No matter what it observed, no agent (whether correct or faulty), can ever rule out the possibility of those observations being artificially manufactured and not real.

If $f \geq 1$, i.e., if <u>at least</u> one agent can become byzantine, no agent can ever know that

- a particular action or event actually happened;
- it itself is correct;
- another agent is byzantine.

If $f \geq 2$, i.e., if <u>more than</u> one agent can become byzantine, no agent can ever know that

- another agent is correct.

This is why knowledge of a trigger event cannot be a precondition!

Brain in a Vat Postulate I
An agent canNOT know its own correctness

Axiom representation

$$\mathit{iByz} := \qquad \neg K_i \neg H_i \bot$$

Frame characterization

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

Brain in a Vat Postulate II
A faulty agent canNOT know whether any other agent is correct or faulty

Axiom representation ($i \neq j$)

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

Frame characterization ($i \neq j$)

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

Logical Derivation of Brain in a Vat

Reminder ($i \neq j$)

$$
\begin{aligned}
iByz &:= \neg K_i \neg H_i \bot \\
BiV &:= H_i \bot \rightarrow \neg K_i H_j \bot \wedge \neg K_i \neg H_j \bot
\end{aligned}
$$

Brain-in-a-Vat Lemma ($i \neq j$)

$$\mathscr{KH} + iByz + BiV \vdash \neg K_i \neg H_j \bot \wedge \neg K_i H_j \bot$$

i.e., no agent knows whether another agent is correct or faulty

Logical Derivation of Brain in a Vat

Reminder ($i \neq j$)

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

Brain-in-a-Vat Lemma ($i \neq j$)

$$\mathscr{KH} + iByz + BiV \vdash \neg K_i \neg H_j \bot \wedge \neg K_i H_j \bot$$

i.e., no agent knows whether another agent is correct or faulty

What about the distinction between $f \geq 1$ and $f \geq 2$?

Distributed systems require at least two faulty agents to prove ignorance about correctness of others.

Logical Explanation of Brain in a Vat

Reminder ($i \neq j$)

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

$$BiV := H_i \bot \rightarrow \neg K_i H_j \bot \wedge \neg K_i \neg H_j \bot$$

$$Byz_1 := \bigvee_{\substack{G \subseteq \mathcal{A} \\ |G|=n-1}} \bigwedge_{j \in G} \neg H_j \bot = \bigvee_{i} \bigwedge_{j \neq i} \neg H_j \bot$$

Brain-in-a-Vat Analysis for $f = 1$ ($i \neq j$)

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

i.e., one conjunct of BiV's conclusion is derivable

Logical Explanation of Brain in a Vat

Reminder ($i \neq j$)

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

$$BiV := H_i \bot \rightarrow \neg K_i H_j \bot \land \neg K_i \neg H_j \bot$$

$$Byz_1 := \bigvee_{\substack{G \subseteq \mathcal{A} \\ |G| = n-1}} \bigwedge_{j \in G} \neg H_j \bot = \bigvee_{i} \bigwedge_{j \neq i} \neg H_j \bot$$

Brain-in-a-Vat Analysis for $f = 1$ ($i \neq j$)

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

i.e., one conjunct of BiV's conclusion is derivable

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

i.e., the other conjunct of BiV is problematic: agents would lose ability to detect own faults

Logical Explanation of Brain in a Vat

Reminder ($i \neq j$)

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

$$BiV := H_i \bot \rightarrow \neg K_i H_j \bot \wedge \neg K_i \neg H_j \bot$$

$$Byz_1 := \bigvee_{\substack{G \subseteq \mathcal{A} \\ |G| = n-1}} \bigwedge_{j \in G} \neg H_j \bot = \bigvee_{i} \bigwedge_{j \neq i} \neg H_j \bot$$

Brain-in-a-Vat Analysis for $f = 1$ ($i \neq j$)

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

i.e., one conjunct of BiV's conclusion is derivable

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

i.e., the other conjunct of BiV is problematic:
agents would lose ability to detect own faults

Logical conclusion

Do not postulate BiV for $f = 1$. Then only $\neg K_i H_j \bot$ remains.

Conclusion

Past Work

- Normal, frame-characterizable logic for byzantine agents
- Completeness theorem
- Completeness with common hope and common knowledge
- Confirmation and explanation of distributed results

Present and Future Work

- Eventual common hope
- Self-stabilizing agents in style of DEL
- A priori knowledge
- Algebraic topological approach (simplicial complexes)
- ...

Thank you!