

When Do Introspection Axioms Matter for Multi-Agent Epistemic Reasoning?

Yifeng Ding, Wesley H. Holliday, Cedegao Zhang

TARK 2019

UC Berkeley

Group of Logic and the Methodology of Science &

Department of Philosophy

Multi-agent epistemic reasoning

- There is a huge literature on introspection axioms.

Multi-agent epistemic reasoning

- There is a huge literature on introspection axioms.
- It is shown that sometimes the introspection axioms are the hidden assumptions behind certain “paradoxical” theorems, like the impossibility of agreeing to disagree.

Multi-agent epistemic reasoning

- There is a huge literature on introspection axioms.
- It is shown that sometimes the introspection axioms are the hidden assumptions behind certain “paradoxical” theorems, like the impossibility of agreeing to disagree.
- Whether it is reasonable to assume them in full is still lively debated in philosophy.

Multi-agent epistemic reasoning

- There is a huge literature on introspection axioms.
- It is shown that sometimes the introspection axioms are the hidden assumptions behind certain “paradoxical” theorems, like the impossibility of agreeing to disagree.
- Whether it is reasonable to assume them in full is still lively debated in philosophy.

However, debates can be tiring.

Multi-agent epistemic reasoning

- There is a huge literature on introspection axioms.
- It is shown that sometimes the introspection axioms are the hidden assumptions behind certain “paradoxical” theorems, like the impossibility of agreeing to disagree.
- Whether it is reasonable to assume them in full is still lively debated in philosophy.

However, debates can be tiring.

- Do we really need to introspect and

Multi-agent epistemic reasoning

- There is a huge literature on introspection axioms.
- It is shown that sometimes the introspection axioms are the hidden assumptions behind certain “paradoxical” theorems, like the impossibility of agreeing to disagree.
- Whether it is reasonable to assume them in full is still lively debated in philosophy.

However, debates can be tiring.

- Do we really need to introspect and
- if we don't even try to introspect, do introspection axioms still matter?

- Formalize “don’t even try to introspect” part.

Plan

- Formalize “don’t even try to introspect” part.
- Formalize “do axioms matter” part.

- Formalize “don’t even try to introspect” part.
- Formalize “do axioms matter” part.
- Give the answer to the formalized if-then question.

- Formalize “don’t even try to introspect” part.
- Formalize “do axioms matter” part.
- Give the answer to the formalized if-then question.
- Provide some details.

- Formalize “don’t even try to introspect” part.
- Formalize “do axioms matter” part.
- Give the answer to the formalized if-then question.
- Provide some details.
- Discuss previous works and possible extensions.

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_1 \wedge \Box_3\neg m_1))$

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_1 \wedge \Box_3\neg m_1))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3\neg m_2))$

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_1 \wedge \Box_3\neg m_1))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3\neg m_2))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3 m_3))$

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_1 \wedge \Box_3\neg m_1))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3\neg m_2))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3 m_3))$
- $\Box_1\Box_2\neg\Box_3 m_3$ (3 didn't step forward in the first round)

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_1 \wedge \Box_3\neg m_1))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3\neg m_2))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3 m_3))$
- $\Box_1\Box_2\neg\Box_3 m_3$ (3 didn't step forward in the first round)
- $\Box_1(\neg m_1 \rightarrow \Box_2 m_2)$

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_1 \wedge \Box_3\neg m_1))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3\neg m_2))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3 m_3))$
- $\Box_1\Box_2\neg\Box_3 m_3$ (3 didn't step forward in the first round)
- $\Box_1(\neg m_1 \rightarrow \Box_2 m_2)$
- $\Box_1\neg\Box_2 m_2$ (2 didn't step forward in the second round)

Introspection in epistemic languages

Observation

In the classic muddy children puzzle, the children don't need to reason about their own beliefs. It can also be formalized such that for any child a , \Box_a never immediately scope over \Box_a .

- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_1 \wedge \Box_3\neg m_1))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3\neg m_2))$
- $\Box_1(\neg m_1 \rightarrow \Box_2(\neg m_2 \rightarrow \Box_3 m_3))$
- $\Box_1\Box_2\neg\Box_3 m_3$ (3 didn't step forward in the first round)
- $\Box_1(\neg m_1 \rightarrow \Box_2 m_2)$
- $\Box_1\neg\Box_2 m_2$ (2 didn't step forward in the second round)
- $\Box_1 m_1$

No introspection in the language

“Don't even try to introspect” = no introspection in the language

Intuition Every agent only thinks about non-modal propositions or other agents.

No introspection in the language

“Don't even try to introspect” = no introspection in the language

Intuition Every agent only thinks about non-modal propositions or other agents.

Formal 1 \Box_a can only scope over a Boolean combination of atomic propositions and formulas of the form $\Box_b\varphi$ with $b \neq a$, and hereditarily so.

No introspection in the language

“Don’t even try to introspect” = no introspection in the language

Intuition Every agent only thinks about non-modal propositions or other agents.

Formal 1 \Box_a can only scope over a Boolean combination of atomic propositions and formulas of the form $\Box_b\varphi$ with $b \neq a$, and hereditarily so.

Formal 2 In the parsing tree restricted to the modal operators, every path from the roots to the leaves is agent-alternating. Modalities of the same agent are never adjacent.

No introspection in the language

“Don’t even try to introspect” = no introspection in the language

Intuition Every agent only thinks about non-modal propositions or other agents.

Formal 1 \Box_a can only scope over a Boolean combination of atomic propositions and formulas of the form $\Box_b\varphi$ with $b \neq a$, and hereditarily so.

Formal 2 In the parsing tree restricted to the modal operators, every path from the roots to the leaves is agent-alternating. Modalities of the same agent are never adjacent.

This idea is not new. We’ll say more about previous works later.

Agent-alternating formulas

Agent-alternating formulas

Define a family $\{\mathcal{L}_{-a}\}_{a \in A}$ of languages through the following simultaneous induction:

$$\mathcal{L}_{-a} \ni \varphi ::= p \mid \Box_x \psi \mid \neg \varphi \mid (\varphi \wedge \varphi)$$

where $p \in \text{Prop}$ and $x \in A \setminus \{a\}$ while $\psi \in \mathcal{L}_{-x}$.

Then the language \mathcal{L}_{alt} is defined inductively by

$$\mathcal{L}_{alt} \ni \varphi ::= p \mid \chi \mid \neg \varphi \mid (\varphi \wedge \varphi)$$

where $p \in \text{Prop}$ and $\chi \in \bigcup_{a \in A} \mathcal{L}_{-a}$.

Agent-alternating formulas

- \mathcal{L}_{-a} is the set of agent-alternating formulas that don't start with \Box_a .

Agent-alternating formulas

- \mathcal{L}_{-a} is the set of agent-alternating formulas that don't start with \Box_a .
- \mathcal{L}_{-a} is the set of Boolean combinations of $\text{Prop} \cup \bigcup_{x \neq a} \Box_x \mathcal{L}_{-x}$.

Agent-alternating formulas

- \mathcal{L}_{-a} is the set of agent-alternating formulas that don't start with \Box_a .
- \mathcal{L}_{-a} is the set of Boolean combinations of $\text{Prop} \cup \bigcup_{x \neq a} \Box_x \mathcal{L}_{-x}$.
- Formulas that don't start with \Box_a have been called “objective formulas for a ”. \mathcal{L}_{-a} is its hereditary version, also studied in the same line of research.

Agent-alternating formulas

- \mathcal{L}_{-a} is the set of agent-alternating formulas that don't start with \Box_a .
- \mathcal{L}_{-a} is the set of Boolean combinations of $\text{Prop} \cup \bigcup_{x \neq a} \Box_x \mathcal{L}_{-x}$.
- Formulas that don't start with \Box_a have been called “objective formulas for a ”. \mathcal{L}_{-a} is its hereditary version, also studied in the same line of research.
- Our notation is intended to mimic its use in game theory.

Agent-alternating formulas

- \mathcal{L}_{-a} is the set of agent-alternating formulas that don't start with \Box_a .
- \mathcal{L}_{-a} is the set of Boolean combinations of $\text{Prop} \cup \bigcup_{x \neq a} \Box_x \mathcal{L}_{-x}$.
- Formulas that don't start with \Box_a have been called “objective formulas for a ”. \mathcal{L}_{-a} is its hereditary version, also studied in the same line of research.
- Our notation is intended to mimic its use in game theory.

$\Box_a(p \wedge \Box_b \Box_a q)$ is agent-alternating. $\Box_a(\Box_b \Box_a p \wedge \Box_a q)$ is not.

Introspection axioms don't matter

“Introspection axioms don't matter” = conservativity.

Introspection axioms don't matter

“Introspection axioms don't matter” = conservativity.

- With or without the introspection axioms, you can make the same inference moves.

Introspection axioms don't matter

“Introspection axioms don't matter” = conservativity.

- With or without the introspection axioms, you can make the same inference moves.
- I.e., the logic (what follows from what) doesn't change.

Introspection axioms don't matter

“Introspection axioms don't matter” = conservativity.

- With or without the introspection axioms, you can make the same inference moves.
- I.e., the logic (what follows from what) doesn't change.
- This means we can do formalization and reasoning with certainty in certain cases while being uncertain about what \square really means and which logic it really follows in full generality.

Question formalized

Hence, the question “if we don’t even try to introspect, do introspection axioms still matter?” is formalized as follows.

Main question

For which modal logic L and which axiom φ ,

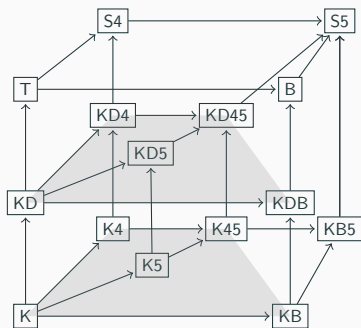
$$L \cap \mathcal{L}_{alt} = L\varphi \cap \mathcal{L}_{alt}?$$

More generally, for which modal logics L and L' ,

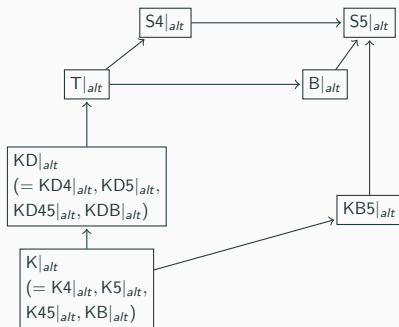
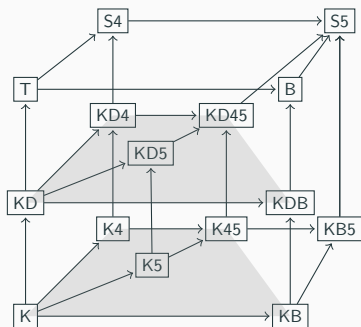
$$L \cap \mathcal{L}_{alt} = L' \cap \mathcal{L}_{alt}?$$

We have a language \mathcal{L}_{alt} , and we ask its power to collapse logics.

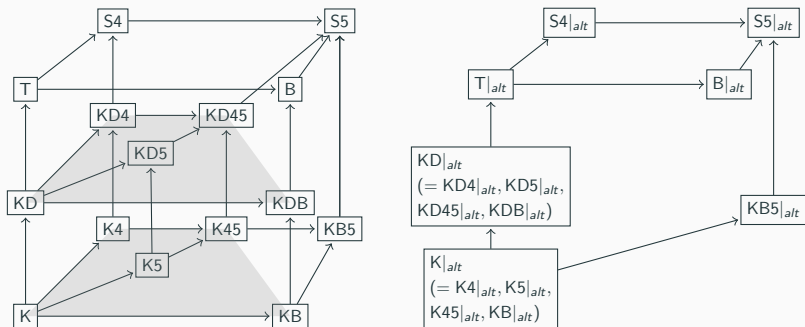
Question answered



Question answered

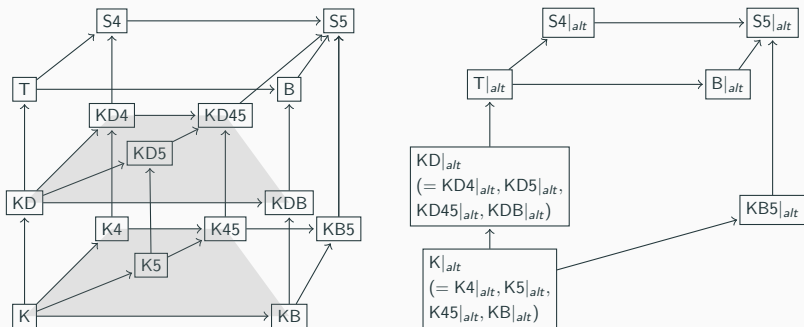


Question answered



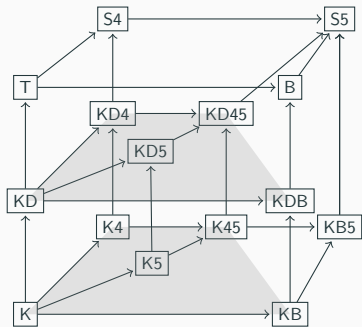
- In particular, $KD|_{alt} = KD45|_{alt}$, but $T|_{alt} \neq S5|_{alt}$.

Question answered

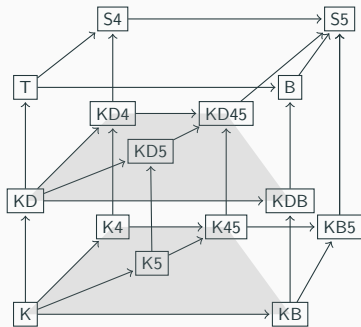


- In particular, $KD|_{alt} = KD45|_{alt}$, but $T|_{alt} \neq S5|_{alt}$.
- $KB5$ almost has T , so no collapse.

Non-collapsing results

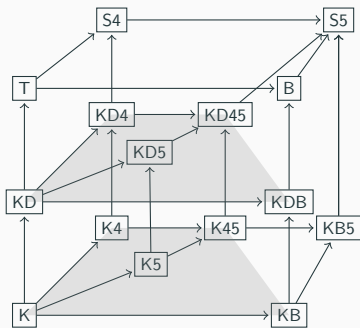


Non-collapsing results



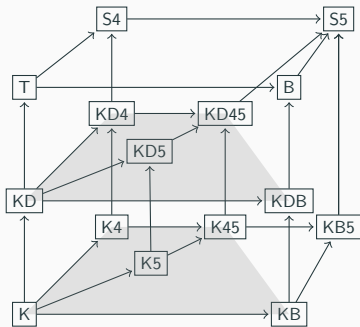
- Non-collapsing of the vertical arrows:
 $T = \Box_a p \rightarrow p$ and $D = \Box_a p \rightarrow \Diamond_a p$
are agent-alternating.

Non-collapsing results



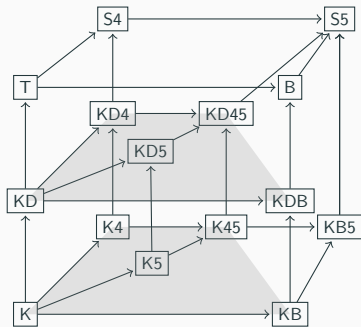
- Non-collapsing of the vertical arrows: $T = \Box_a p \rightarrow p$ and $D = \Box_a p \rightarrow \Diamond_a p$ are agent-alternating.
- On the top layer: we can pad $\Diamond_a \Diamond_a$ with a \Box_b .

Non-collapsing results



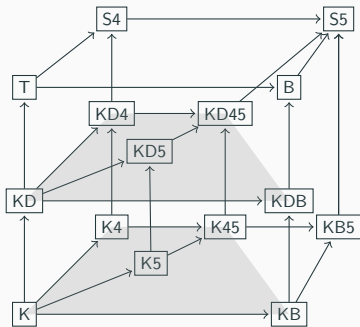
- Non-collapsing of the vertical arrows: $T = \Box_a p \rightarrow p$ and $D = \Box_a p \rightarrow \Diamond_a p$ are agent-alternating.
- On the top layer: we can pad $\Diamond_a \Diamond_a$ with a \Box_b .
- $\Diamond_a \Box_b \Diamond_a p \rightarrow \Diamond_a p$ is agent-alternating, in S4 but not in B.

Non-collapsing results



- Non-collapsing of the vertical arrows: $T = \Box_a p \rightarrow p$ and $D = \Box_a p \rightarrow \Diamond_a p$ are agent-alternating.
- On the top layer: we can pad $\Diamond_a \Diamond_a$ with a \Box_b .
- $\Diamond_a \Box_b \Diamond_a p \rightarrow \Diamond_a p$ is agent-alternating, in S4 but not in B.
- $\Diamond_a \Box_b \Box_a p \rightarrow p \in B|_{alt} \setminus S4|_{alt}$

Non-collapsing results



- Non-collapsing of the vertical arrows:
 $T = \Box_a p \rightarrow p$ and $D = \Box_a p \rightarrow \Diamond_a p$
 are agent-alternating.
- On the top layer: we can pad $\Diamond_a \Diamond_a$
 with a \Box_b .
- $\Diamond_a \Box_b \Diamond_a p \rightarrow \Diamond_a p$ is agent-alternating,
 in S4 but not in B.
- $\Diamond_a \Box_b \Box_a p \rightarrow p \in B|_{alt} \setminus S4|_{alt}$
- For KB5, we add $\Diamond_a \Diamond_b \top$ in the antecedent since in KB5 this
 guarantees that \Box_b is factive.

Collapsing results

Let's show that $K45|_{alt} \subseteq K|_{alt}$ and $KD45|_{alt} \subseteq KD|_{alt}$.

Collapsing results

Let's show that $K45|_{alt} \subseteq K|_{alt}$ and $KD45|_{alt} \subseteq KD|_{alt}$.

- If $\varphi \in \mathcal{L}_{alt}$ has a countermodel, then it has a T&E countermodel. If it has a S countermodel, then it has a ST&E countermodel.

Collapsing results

Let's show that $K45|_{alt} \subseteq K|_{alt}$ and $KD45|_{alt} \subseteq KD|_{alt}$.

- If $\varphi \in \mathcal{L}_{alt}$ has a countermodel, then it has a T&E countermodel. If it has a S countermodel, then it has a ST&E countermodel.
- All models can be transitivised and Euclideanized, preserving seriality and truths in \mathcal{L}_{alt} .

Collapsing results

Let's show that $K45|_{alt} \subseteq K|_{alt}$ and $KD45|_{alt} \subseteq KD|_{alt}$.

- If $\varphi \in \mathcal{L}_{alt}$ has a countermodel, then it has a T&E countermodel. If it has a S countermodel, then it has a ST&E countermodel.
- All models can be transitivised and Euclideanized, preserving seriality and truths in \mathcal{L}_{alt} .
- Agent-alternating bisimulation family:
a \leftrightarrow_{-a} for each \mathcal{L}_{-a} and a \leftrightarrow_{alt} for \mathcal{L}_{alt} , interacting correctly.

Collapsing results

Let's show that $K45|_{alt} \subseteq K|_{alt}$ and $KD45|_{alt} \subseteq KD|_{alt}$.

- If $\varphi \in \mathcal{L}_{alt}$ has a countermodel, then it has a T&E countermodel. If it has a S countermodel, then it has a ST&E countermodel.
- All models can be transitivised and Euclideanized, preserving seriality and truths in \mathcal{L}_{alt} .
- Agent-alternating bisimulation family:
a \leftrightarrow_{-a} for each \mathcal{L}_{-a} and a \leftrightarrow_{alt} for \mathcal{L}_{alt} , interacting correctly.
- Agent-alternating unravelling:
keep only agent-alternating paths so that transitivity is trivial.

Collapsing results

Let's show that $K45|_{alt} \subseteq K|_{alt}$ and $KD45|_{alt} \subseteq KD|_{alt}$.

- If $\varphi \in \mathcal{L}_{alt}$ has a countermodel, then it has a T&E countermodel. If it has a S countermodel, then it has a ST&E countermodel.
- All models can be transitivised and Euclideanized, preserving seriality and truths in \mathcal{L}_{alt} .
- Agent-alternating bisimulation family:
a \leftrightarrow_{-a} for each \mathcal{L}_{-a} and a \leftrightarrow_{alt} for \mathcal{L}_{alt} , interacting correctly.
- Agent-alternating unravelling:
keep only agent-alternating paths so that transitivity is trivial.
- Once unravelled agent-alternatingly, we can add arrows and still be agent-alternatingly bisimilar.

- We are not asking expressivity questions for \mathcal{L}_{alt} *per se* in the traditional way, but its bisimulation shows that it is not very expressive in the right way to collapse a lot of logics.

Side note

- We are not asking expressivity questions for \mathcal{L}_{alt} *per se* in the traditional way, but its bisimulation shows that it is not very expressive in the right way to collapse a lot logics.
- As it happens, in K45 (and hence KD45), \mathcal{L} is no more expressive than \mathcal{L}_{alt} .

- We are not asking expressivity questions for \mathcal{L}_{alt} *per se* in the traditional way, but its bisimulation shows that it is not very expressive in the right way to collapse a lot logics.
- As it happens, in K45 (and hence KD45), \mathcal{L} is no more expressive than \mathcal{L}_{alt} .
 - So above K45, no collapse!

- We are not asking expressivity questions for \mathcal{L}_{alt} *per se* in the traditional way, but its bisimulation shows that it is not very expressive in the right way to collapse a lot logics.
- As it happens, in K45 (and hence KD45), \mathcal{L} is no more expressive than \mathcal{L}_{alt} .
 - So above K45, no collapse!
 - And \mathcal{L}_{alt} is not collapsing 45 trivially. It says all that can be said (in \mathcal{L}_{alt}) among T and E models.

Side note

- We are not asking expressivity questions for \mathcal{L}_{alt} *per se* in the traditional way, but its bisimulation shows that it is not very expressive in the right way to collapse a lot logics.
- As it happens, in K45 (and hence KD45), \mathcal{L} is no more expressive than \mathcal{L}_{alt} .
 - So above K45, no collapse!
 - And \mathcal{L}_{alt} is not collapsing 45 trivially. It says all that can be said (in \mathcal{L}_{alt}) among T and E models.
- We also showed that 4 and 5 are necessary among the logics in the Cube for \mathcal{L} and $\mathcal{L}|_{alt}$ to be equi-expressive.

Previous works

The idea of agent-alternating formulas appeared in different places.

Previous works

The idea of agent-alternating formulas appeared in different places.

- In epistemic planning, \mathcal{L}_{alt} is used for efficient reasoning in \mathcal{L} under K45. In fact, $K45|_{alt} = K|_{alt}$ was stated very early (Halpern, Lakemeyer, Shore), though we are unable to locate an explicit proof.

The idea of agent-alternating formulas appeared in different places.

- In epistemic planning, \mathcal{L}_{alt} is used for efficient reasoning in \mathcal{L} under K45. In fact, $K45|_{alt} = K|_{alt}$ was stated very early (Halpern, Lakemeyer, Shore), though we are unable to locate an explicit proof.
- In refinement quantification logics, \mathcal{L}_{alt} is used for axiomatization.

Previous works

The idea of agent-alternating formulas appeared in different places.

- In epistemic planning, \mathcal{L}_{alt} is used for efficient reasoning in \mathcal{L} under K45. In fact, $K45|_{alt} = K|_{alt}$ was stated very early (Halpern, Lakemeyer, Shore), though we are unable to locate an explicit proof.
- In refinement quantification logics, \mathcal{L}_{alt} is used for axiomatization.
- The idea of agent-alternating is very prominent in Bernheim's one of the first papers defining rationalizable strategies, resulting in an agent-alternating system of beliefs.

Previous works

The idea of agent-alternating formulas appeared in different places.

- In epistemic planning, \mathcal{L}_{alt} is used for efficient reasoning in \mathcal{L} under K45. In fact, $K45|_{alt} = K|_{alt}$ was stated very early (Halpern, Lakemeyer, Shore), though we are unable to locate an explicit proof.
- In refinement quantification logics, \mathcal{L}_{alt} is used for axiomatization.
- The idea of agent-alternating is very prominent in Bernheim's one of the first papers defining rationalizable strategies, resulting in an agent-alternating system of beliefs.
 - In fact, we can formalize and prove using Kripke models that agent-alternating common belief of rationality implies the played strategy is rationalizable.

- We can collapse S5 to T if \Box_a is never allowed in the scope of \Box_a . We call these formulas agent-nonrepeating.

Extensions

- We can collapse S5 to T if \Box_a is never allowed in the scope of \Box_a . We call these formulas agent-nonrepeating.
- We can add the usual common knowledge operator and see if there's a natural fragment in line with the idea of “agent alternating” that collapse logics.

Extensions

- We can collapse S5 to T if \Box_a is never allowed in the scope of \Box_a . We call these formulas agent-nonrepeating.
- We can add the usual common knowledge operator and see if there's a natural fragment in line with the idea of “agent alternating” that collapse logics.
- We only have non-collapsing results now. The usual common knowledge is by itself not agent-alternating...

$$\left(\bigwedge_{b \in A \setminus \{a\}} (\Box_b p \wedge C \Box_b p \wedge \Box_b \Box_a p \wedge C \Box_b \Box_a p) \wedge \Box_a p \right) \rightarrow C p.$$

is valid with transitivity, but not otherwise. There should be collapse with infinitely many agents.

Future work

- Collapsing results for natural fragments with the standard common knowledge.

Future work

- Collapsing results for natural fragments with the standard common knowledge.
- A non-trivial fragment collapsing S5 to T.

Future work

- Collapsing results for natural fragments with the standard common knowledge.
- A non-trivial fragment collapsing S5 to T.
- Algebraically, adding axioms is quotienting Lindenbaum algebras. Adding axioms doesn't matter can then be characterized algebraically as a subalgebra is invariant under a quotient. What can we do from the algebraic perspective?

Future work

- Collapsing results for natural fragments with the standard common knowledge.
- A non-trivial fragment collapsing S5 to T.
- Algebraically, adding axioms is quotienting Lindenbaum algebras. Adding axioms doesn't matter can then be characterized algebraically as a subalgebra is invariant under a quotient. What can we do from the algebraic perspective?
- Same question for epistemic logics in richer languages. For example, which formulas with public announcements are agent alternating?

Future work

- Collapsing results for natural fragments with the standard common knowledge.
- A non-trivial fragment collapsing S5 to T.
- Algebraically, adding axioms is quotienting Lindenbaum algebras. Adding axioms doesn't matter can then be characterized algebraically as a subalgebra is invariant under a quotient. What can we do from the algebraic perspective?
- Same question for epistemic logics in richer languages. For example, which formulas with public announcements are agent alternating?
- Finally, can we say more about the practical sufficiency of \mathcal{L}_{alt} ?

Thank you!