

DEL Essentials

Rustam Galimullin

`rustam.galimullin@uib.no`

University of Bergen, Norway

VvL Essentials



Dynamic Epistemic Logic

Plan of the Talk

Part I. Epistemic Logic

Part II. Public Announcement Logic

Part III. Action Models

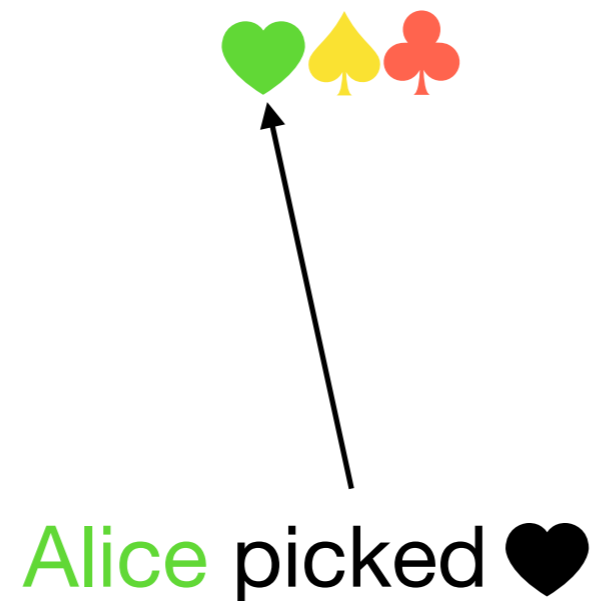
Part IV. Current Research Directions

Part I

Epistemic Logic

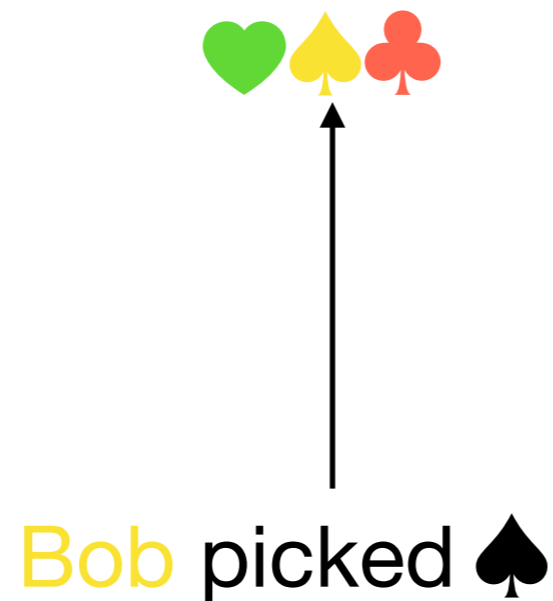
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



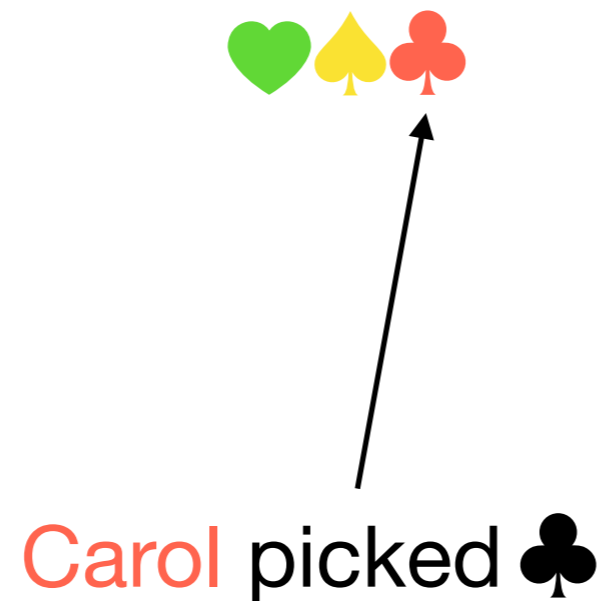
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



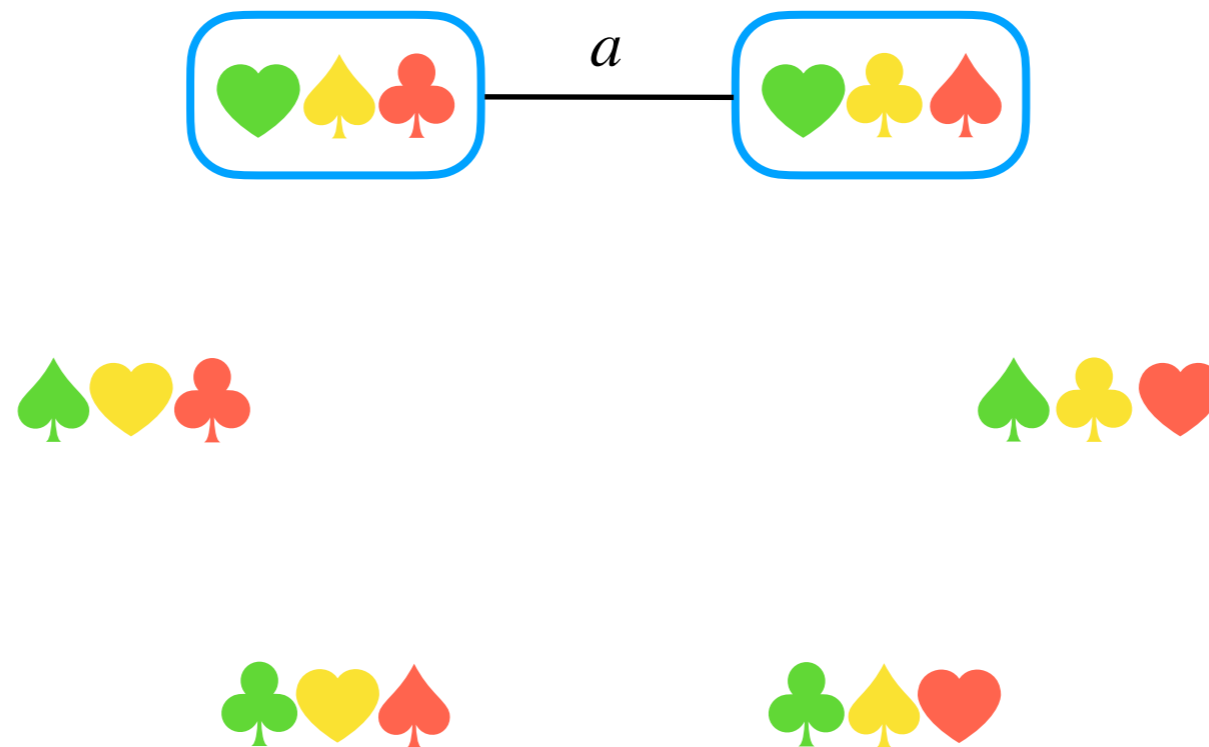
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



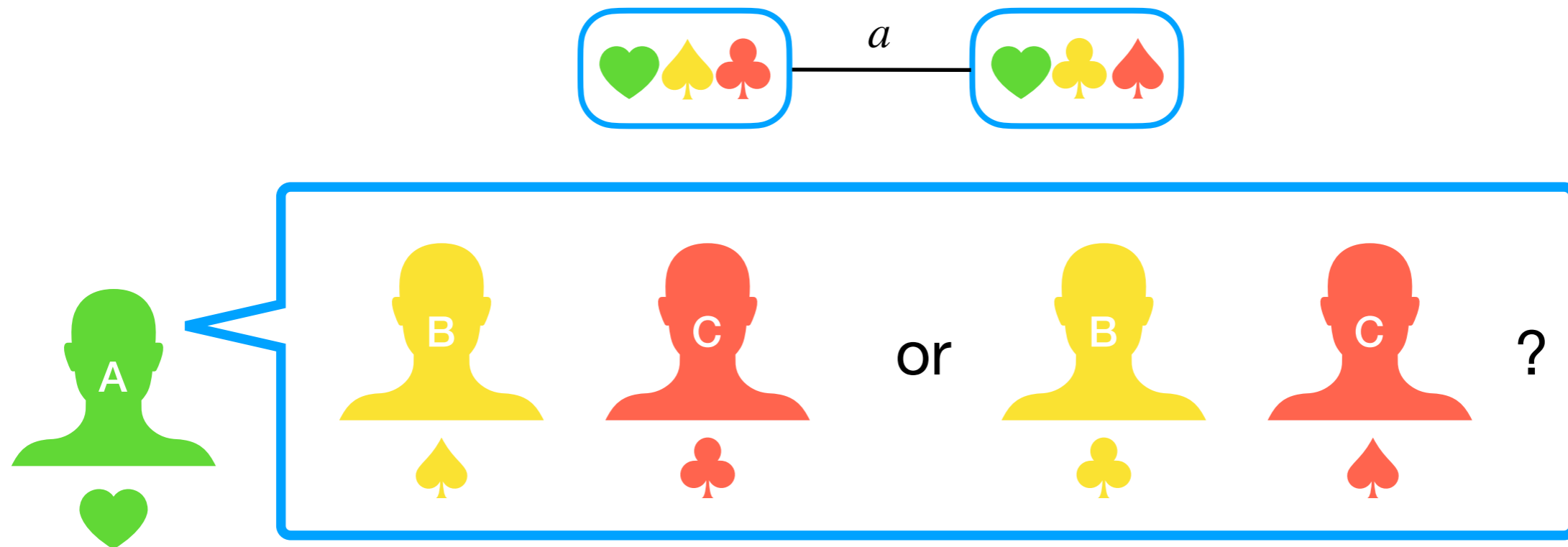
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



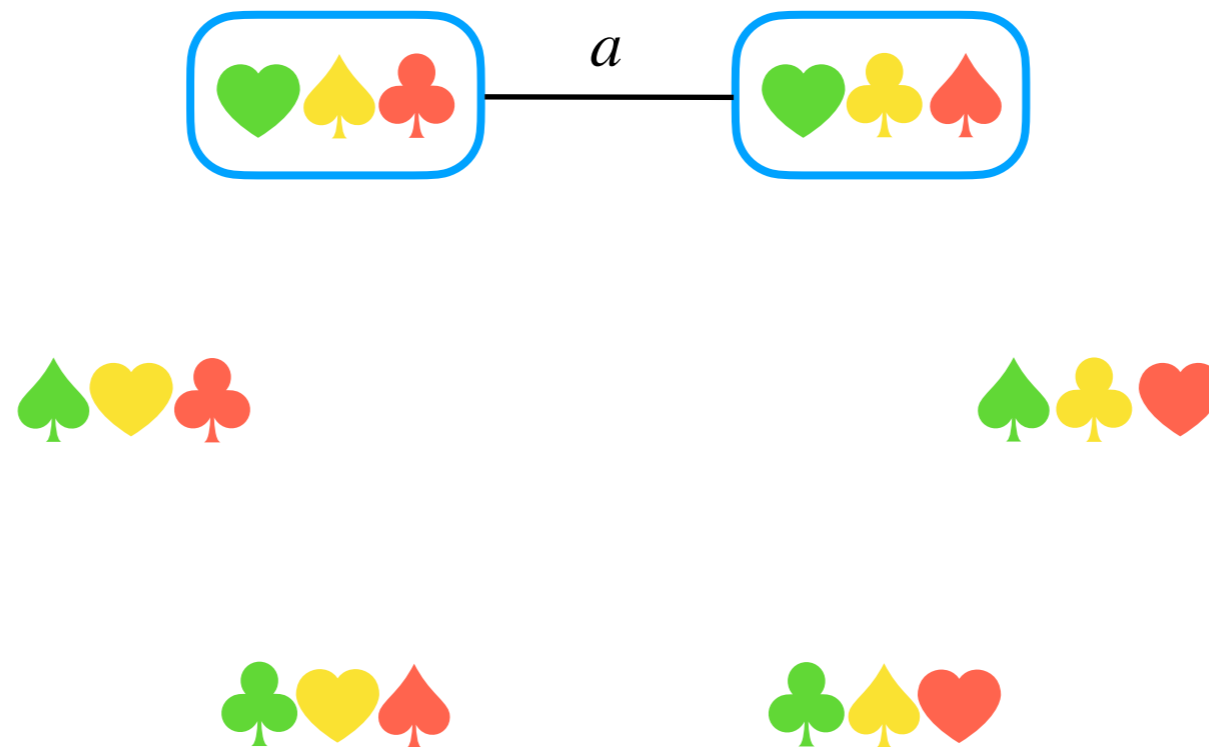
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



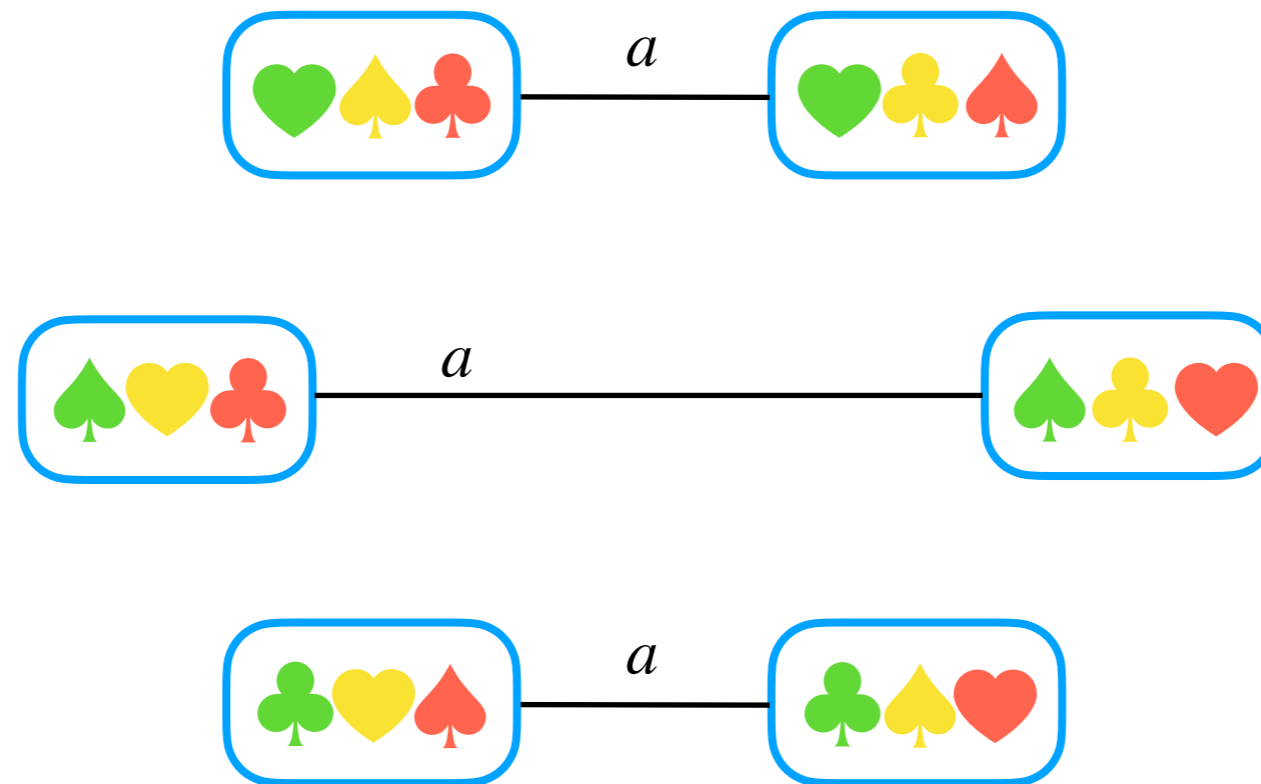
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



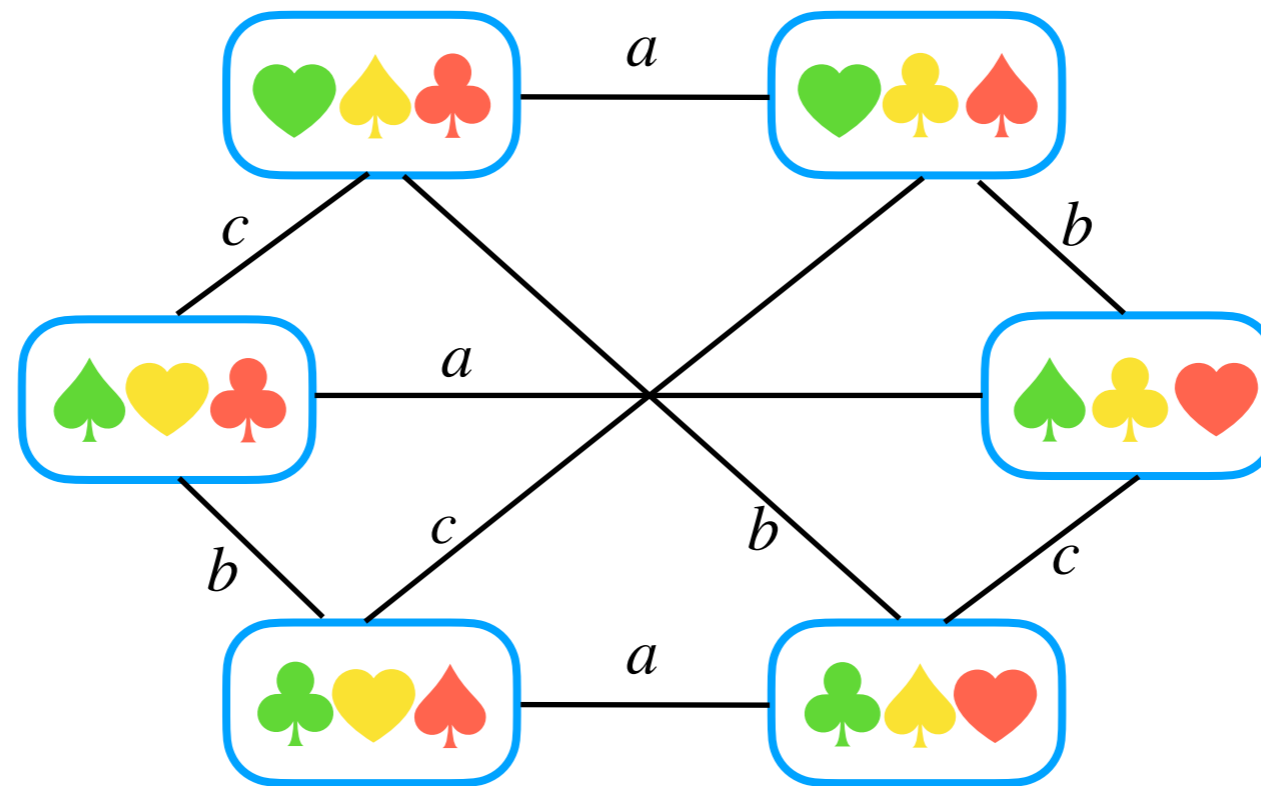
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



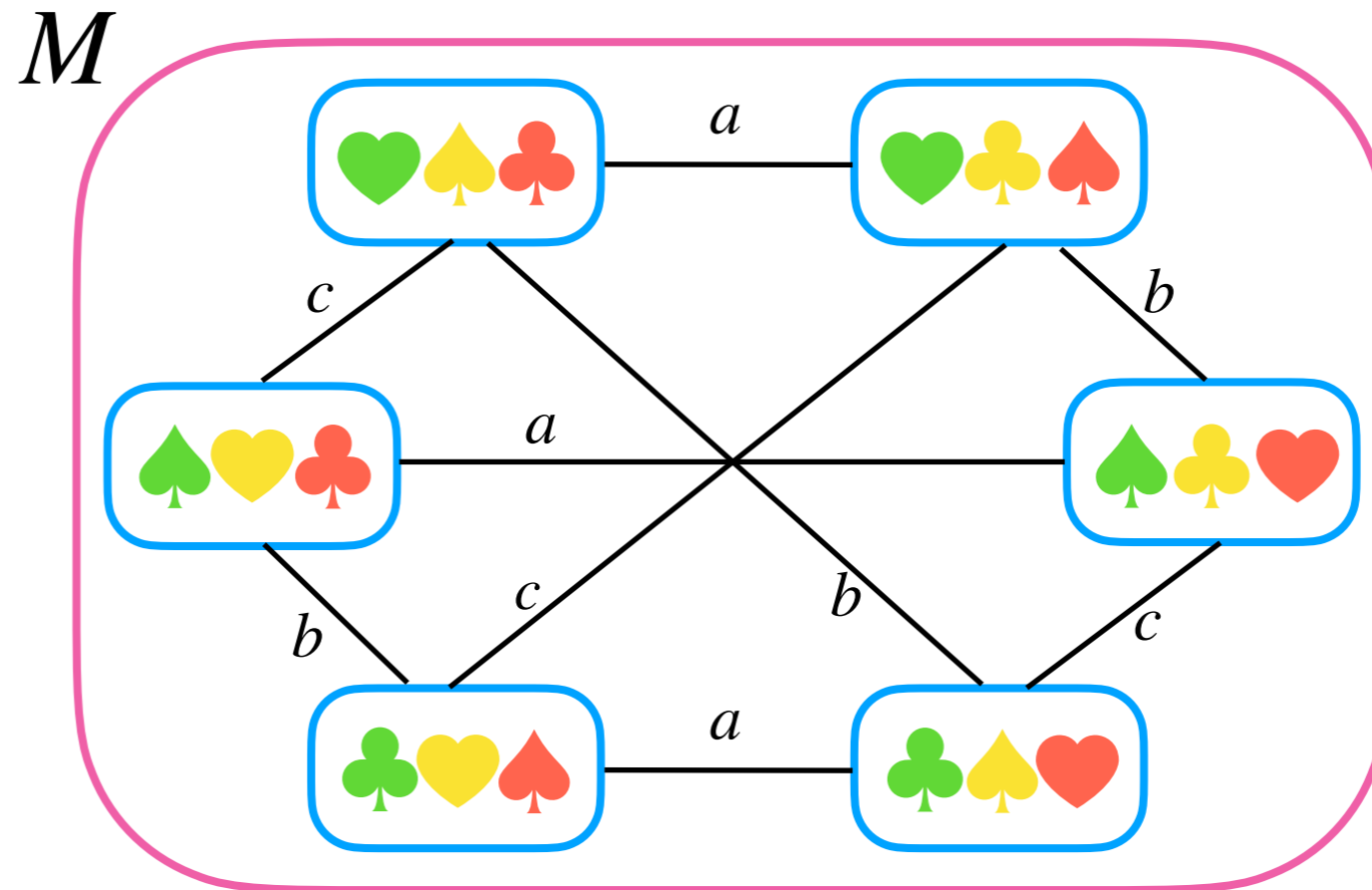
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



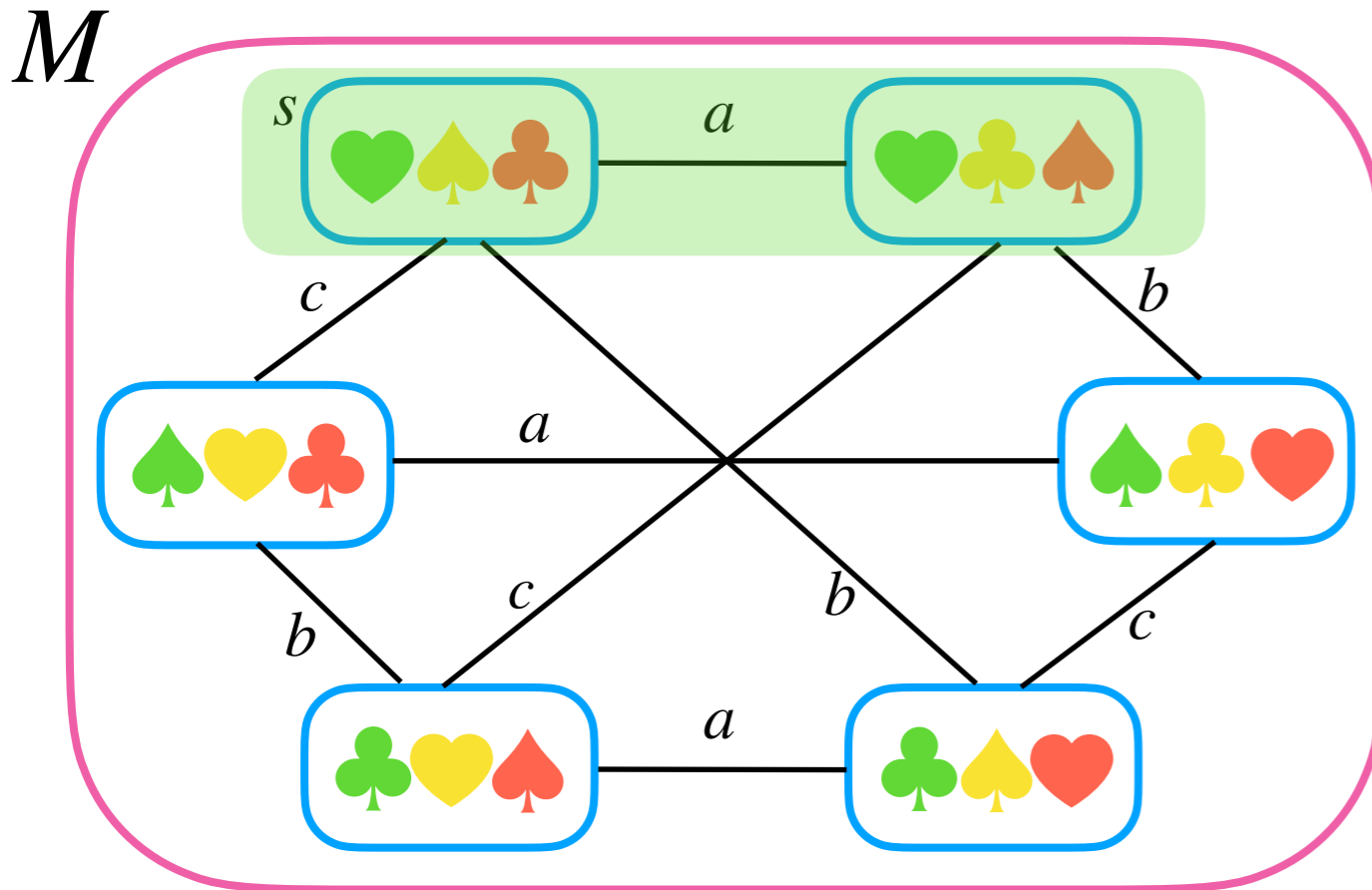
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



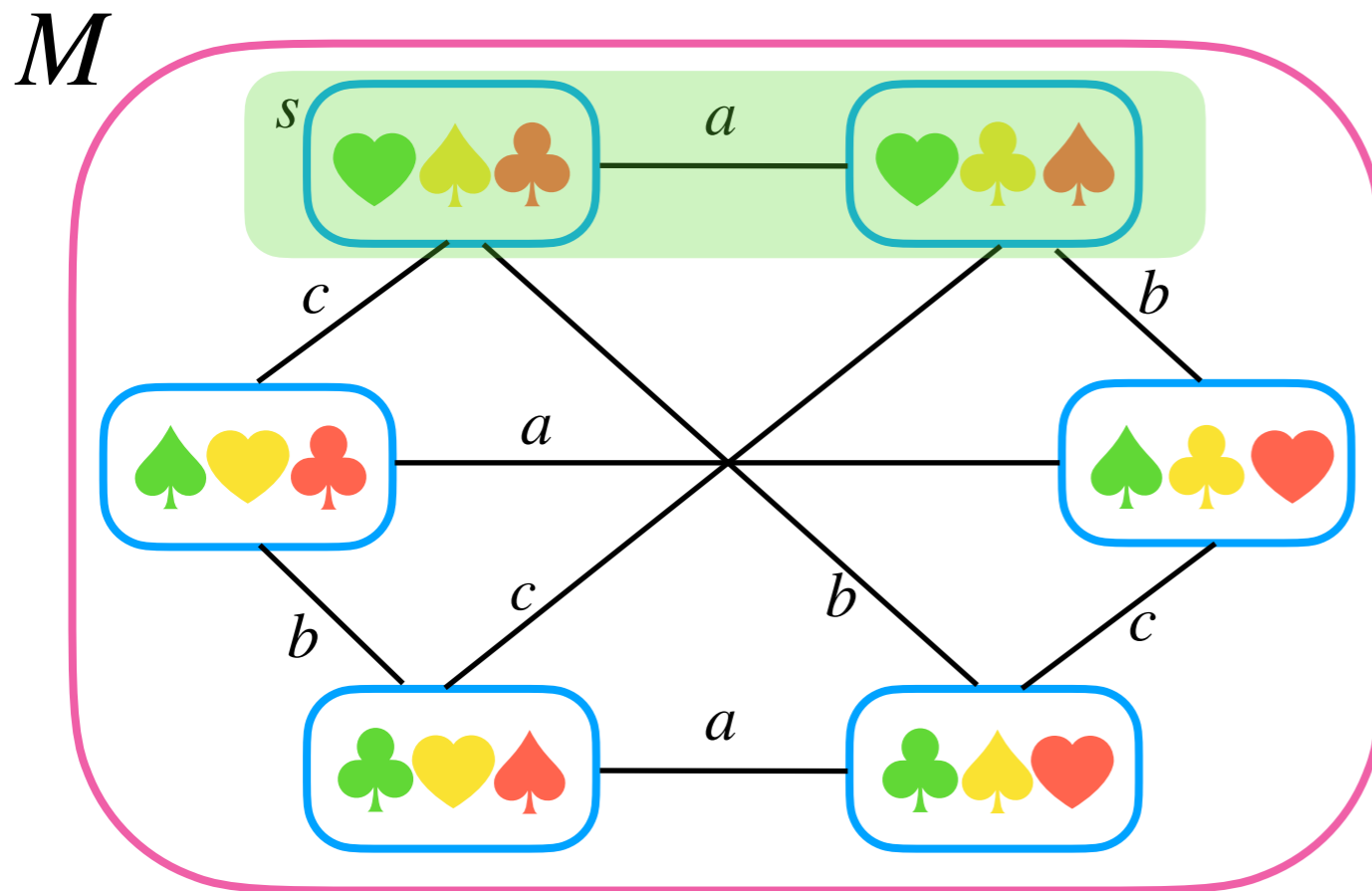
$$M_s \models \heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c$$

$$M_s \models \Box_a (\spadesuit_b \vee \clubsuit_b)$$

$\Box_a \varphi$: An agent **a** **knows** φ if φ is true in all a -reachable states

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



$$M_s \models \heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c$$

$$M_s \models \Box_a (\spadesuit_b \vee \clubsuit_b)$$

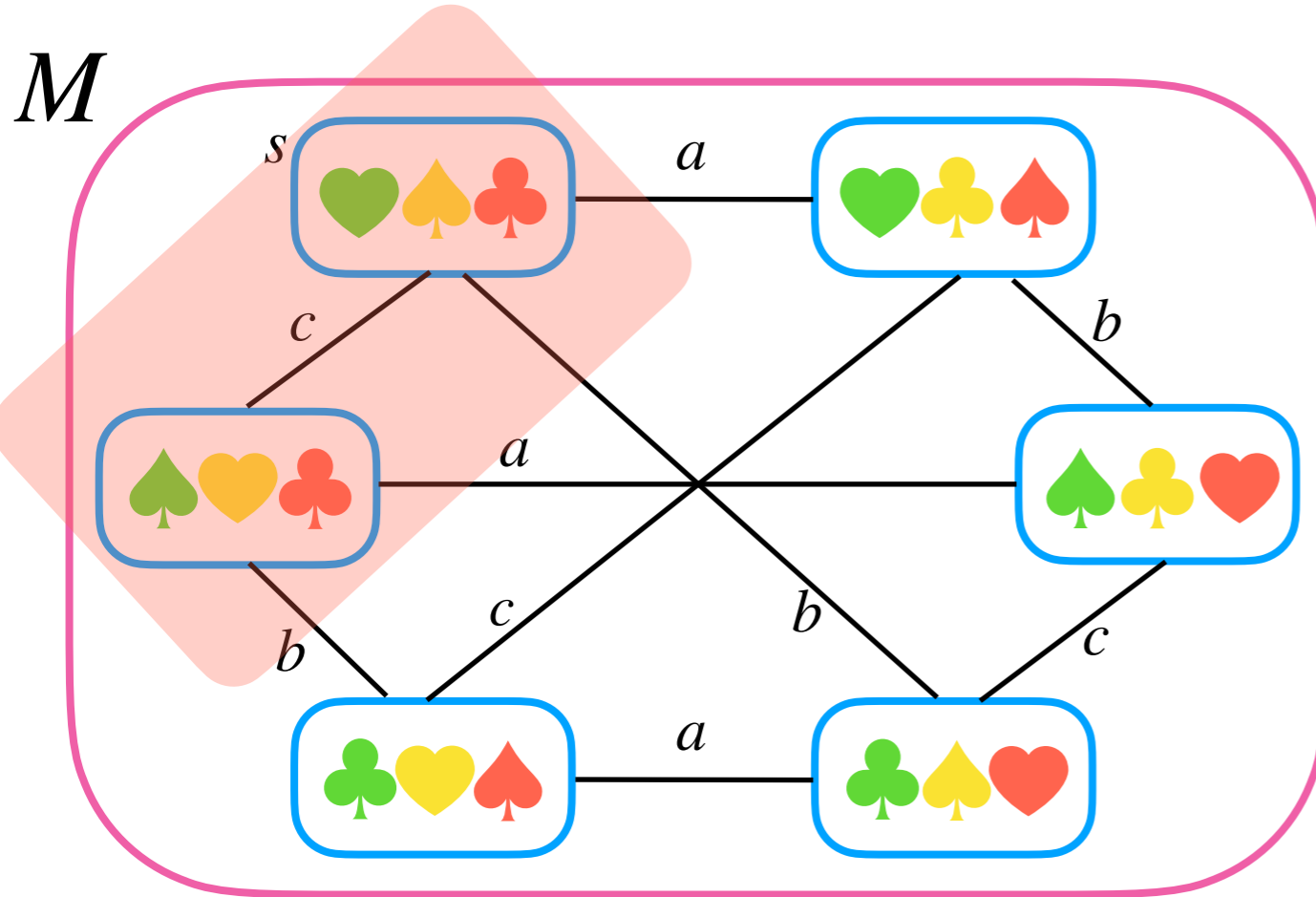
$$M_s \models \Diamond_a (\spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models \Box_c \Box_b \clubsuit_c$$

$\Diamond_a \varphi$: An agent a considers φ possible if φ is true in at least one a -reachable state

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



$$M_s \models \heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c$$

$$M_s \models \Box_a (\spadesuit_b \vee \clubsuit_b)$$

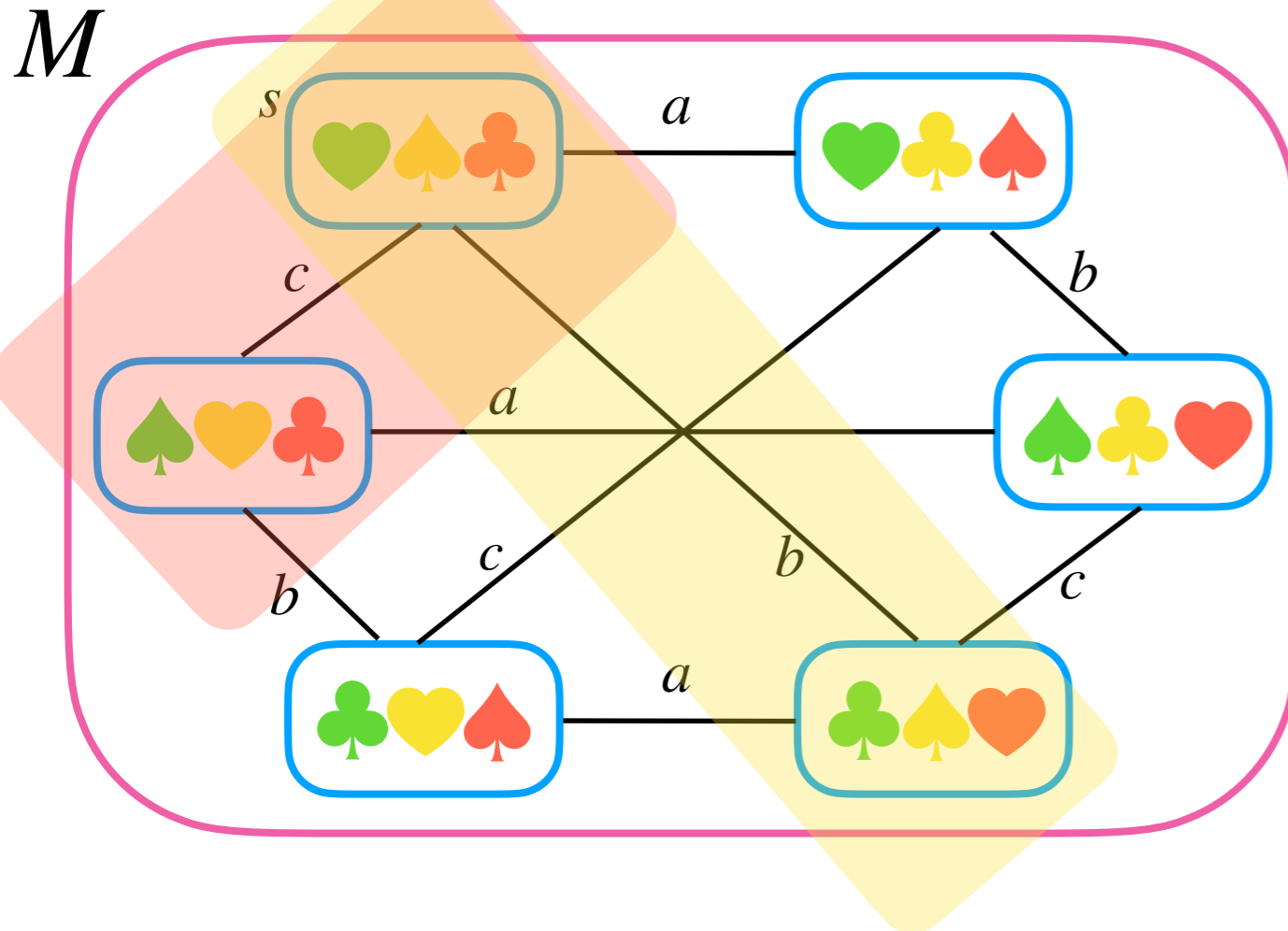
$$M_s \models \Diamond_a (\spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models \Box_c \Box_b \clubsuit_c$$

$\Box_a \varphi$: An agent **a** knows φ if φ is true in all a -reachable states

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



$$M_s \models \heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c$$

$$M_s \models \Box_a (\spadesuit_b \vee \clubsuit_b)$$

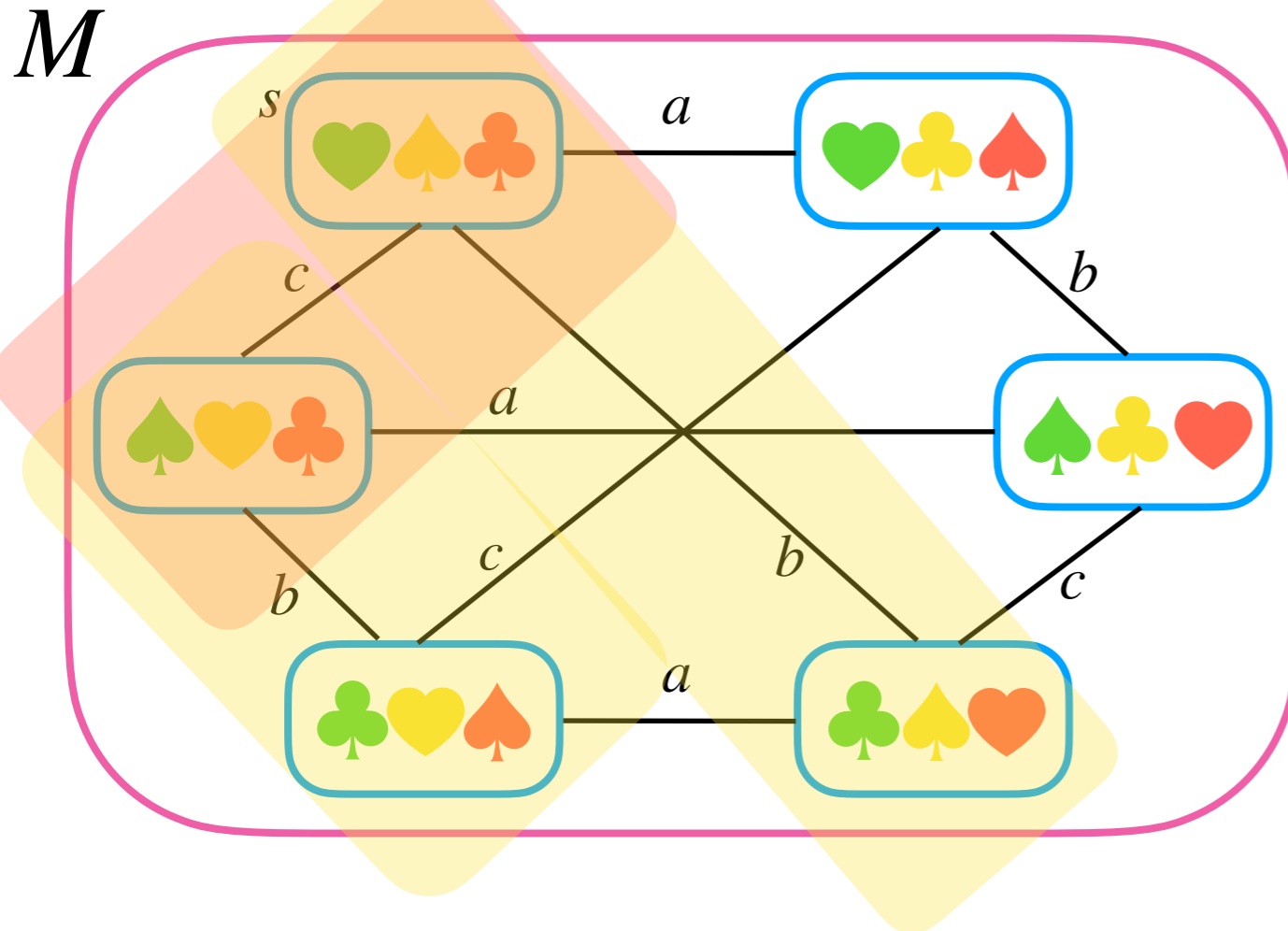
$$M_s \models \Diamond_a (\spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models \Box_c \Box_b \clubsuit_c$$

$\Box_a \varphi$: An agent **a** knows φ if φ is true in all a -reachable states

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



$$M_s \models \heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c$$

$$M_s \models \Box_a (\spadesuit_b \vee \clubsuit_b)$$

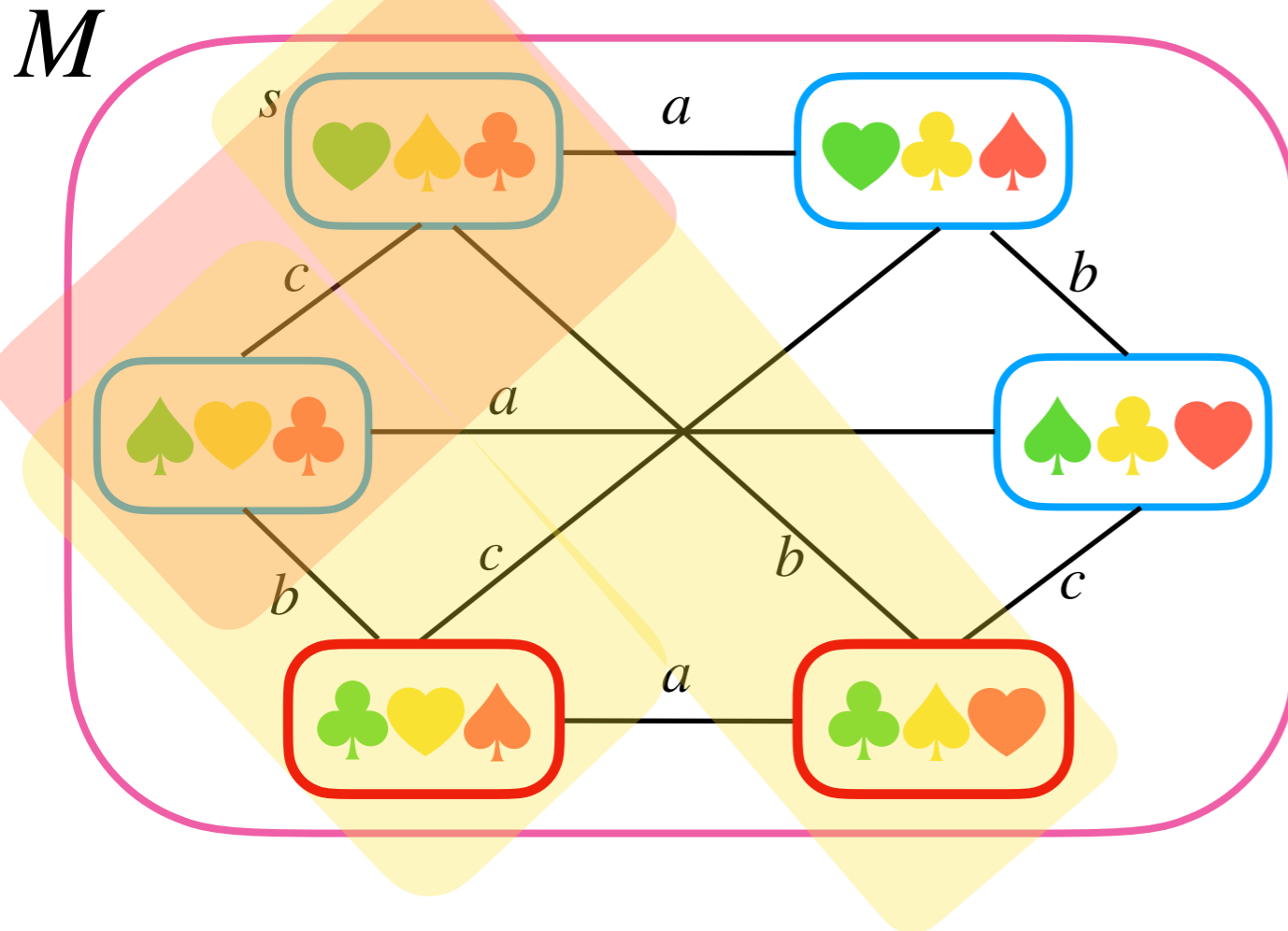
$$M_s \models \Diamond_a (\spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models \Box_c \Box_b \clubsuit_c$$

$\Box_a \varphi$: An agent a knows φ if φ is true in all a -reachable states

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}



$$M_s \models \heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c$$

$$M_s \models \Box_a (\spadesuit_b \vee \clubsuit_b)$$

$$M_s \models \Diamond_a (\spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models \Box_c \Box_b \clubsuit_c$$

$\Box_a \varphi$: An agent **a** **knows** φ if φ is true in all **a**-reachable states

Epistemic Logic

From Greek *episteme* that means *knowledge*

Language of EL $\mathcal{EL} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \Box_a \varphi$

p

$p \wedge q$

$\neg(p \wedge q)$

$\Box_a \neg(p \wedge q)$

$\neg \Box_b \Box_a \neg(p \wedge q)$

Epistemic Logic

Language of EL $\mathcal{EL} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi$

- Epistemic models
- An **epistemic model** M is a tuple (S, \sim, V) , where
- $S \neq \emptyset$ is a set of states;
 - $\sim: A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an equivalence relation;
 - $V: P \rightarrow 2^S$ is the valuation function.

Epistemic Logic

Epistemic
models

An **epistemic model** M is a tuple (S, \sim, V) , where

- $S \neq \emptyset$ is a set of states;
- $\sim: A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an equivalence relation;
- $V: P \rightarrow 2^S$ is the valuation function.

M

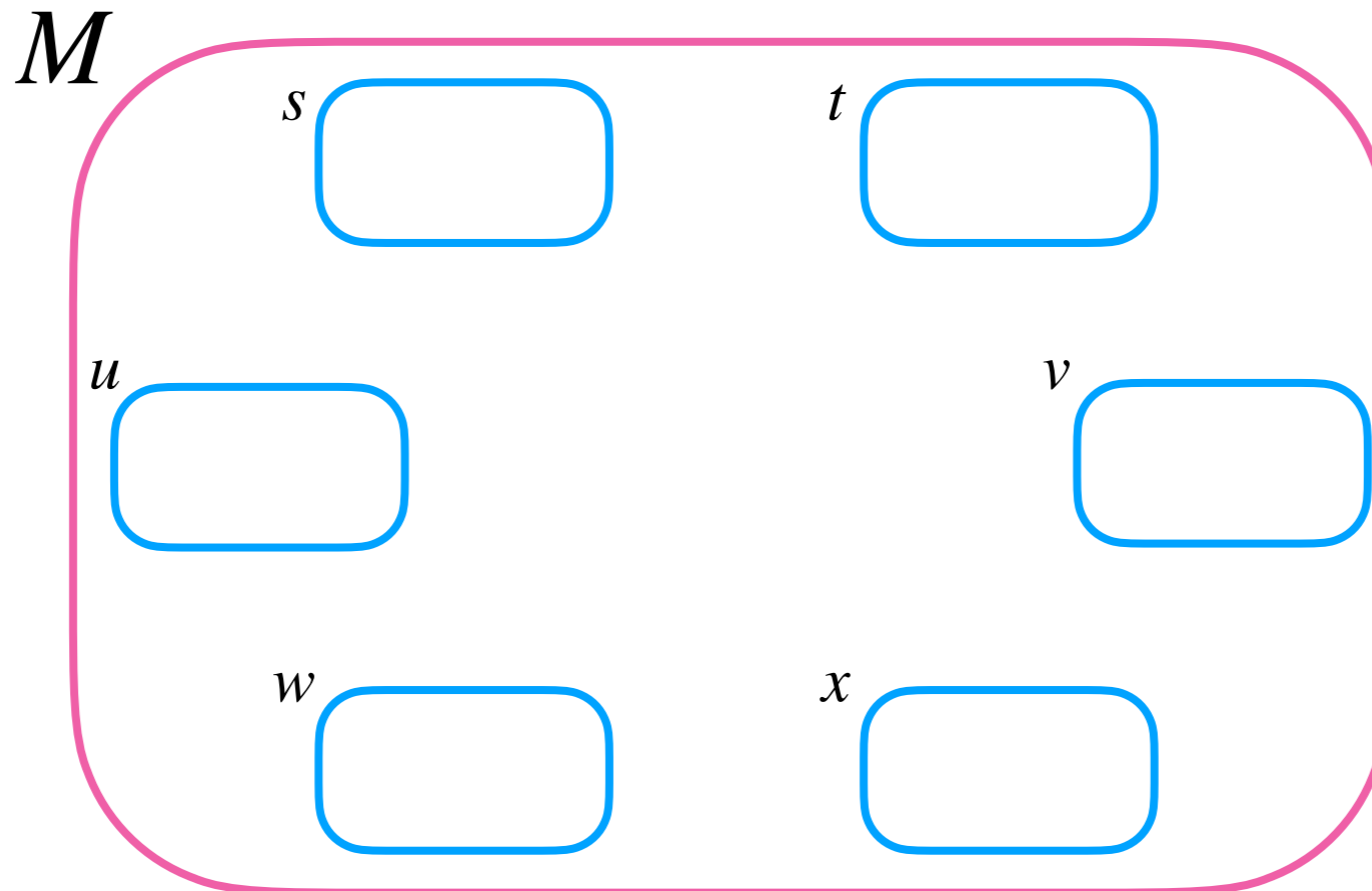


Epistemic Logic

Epistemic models

An **epistemic model** M is a tuple (S, \sim, V) , where

- $S \neq \emptyset$ is a set of states;
- $\sim: A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an equivalence relation;
- $V: P \rightarrow 2^S$ is the valuation function.

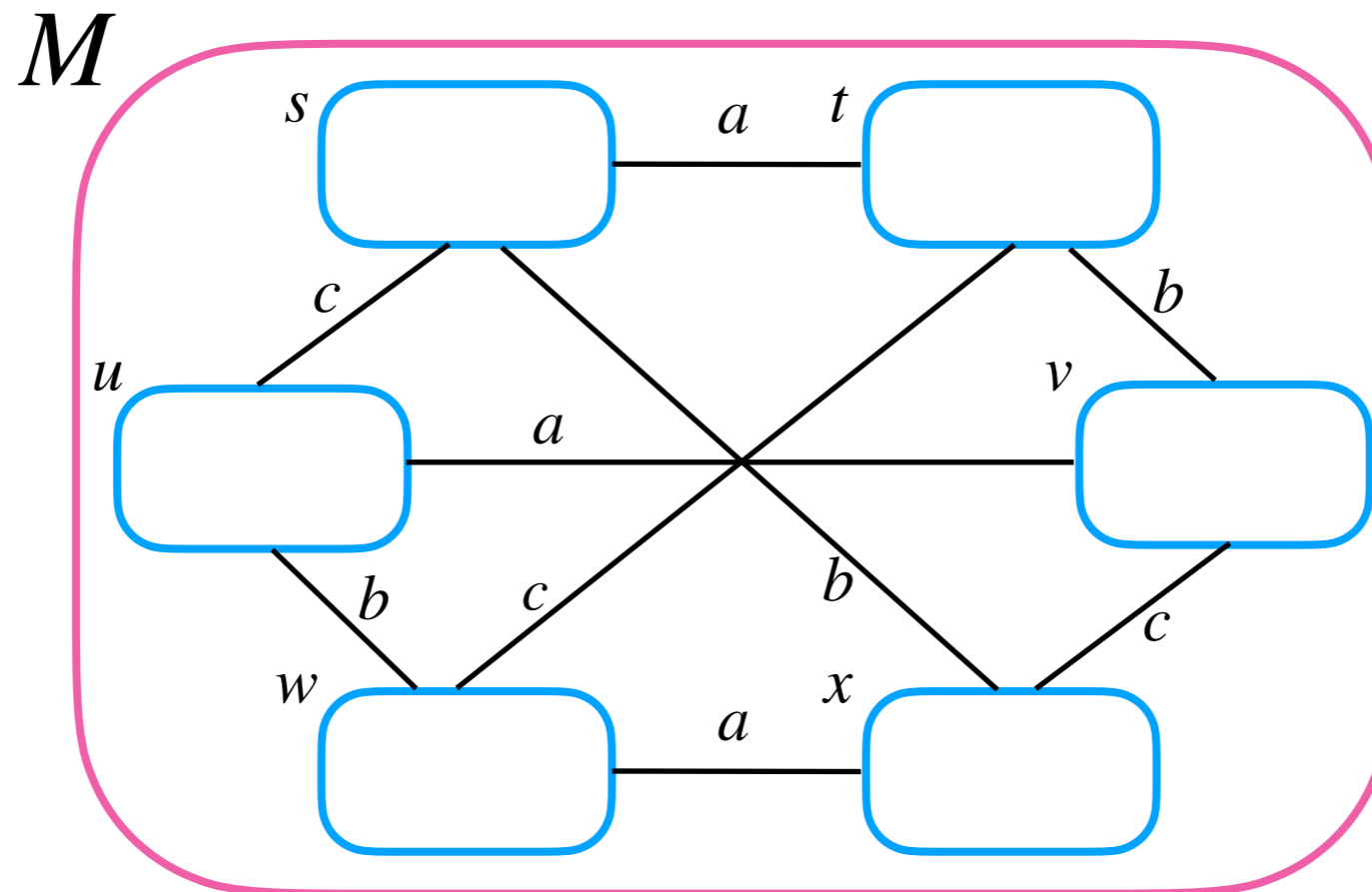


Epistemic Logic

Epistemic models

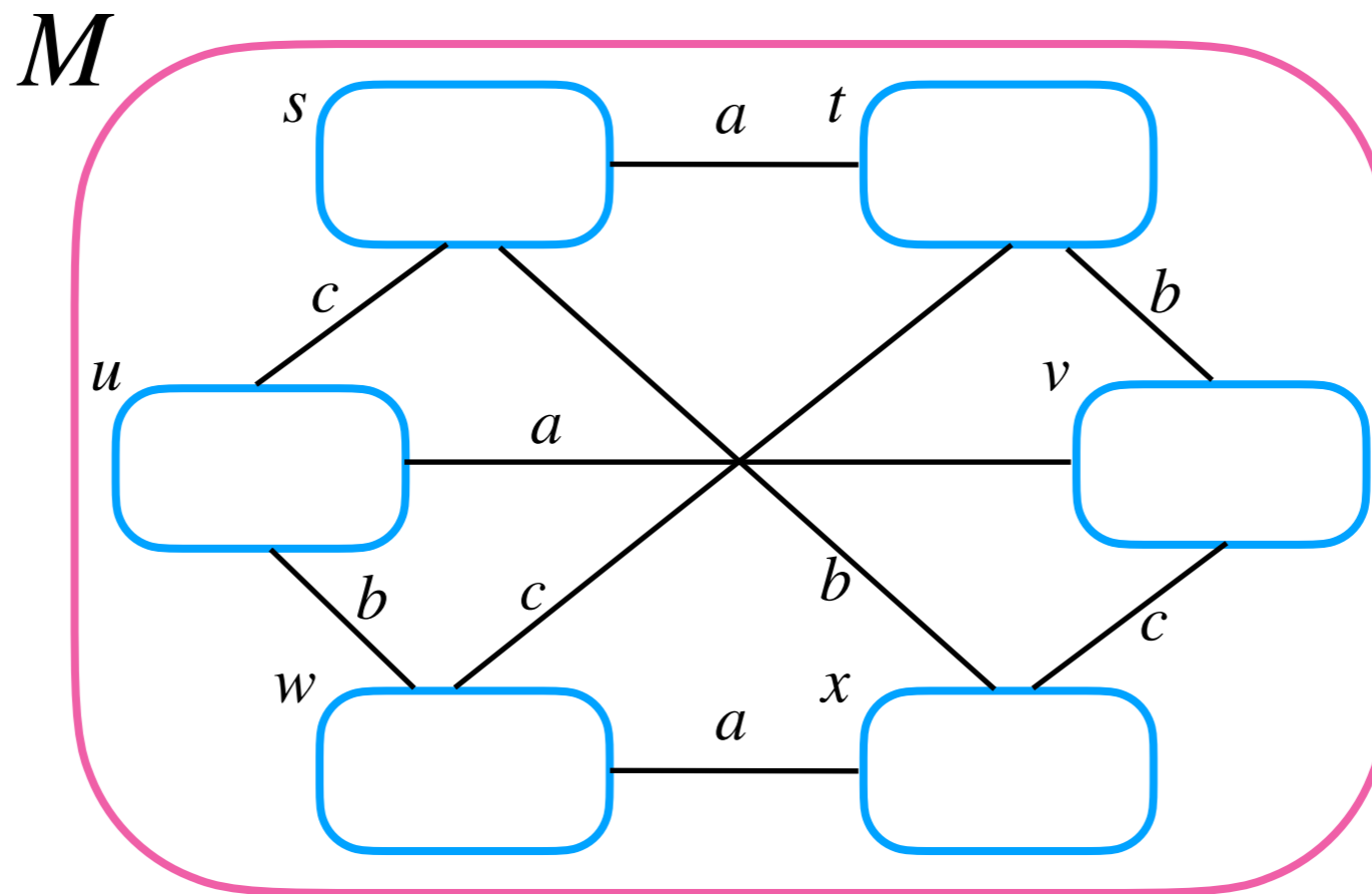
An **epistemic model** M is a tuple (S, \sim, V) , where

- $S \neq \emptyset$ is a set of states;
- $\sim: A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an **equivalence relation**;
- $V: P \rightarrow 2^S$ is the valuation function.



A Quick Aside

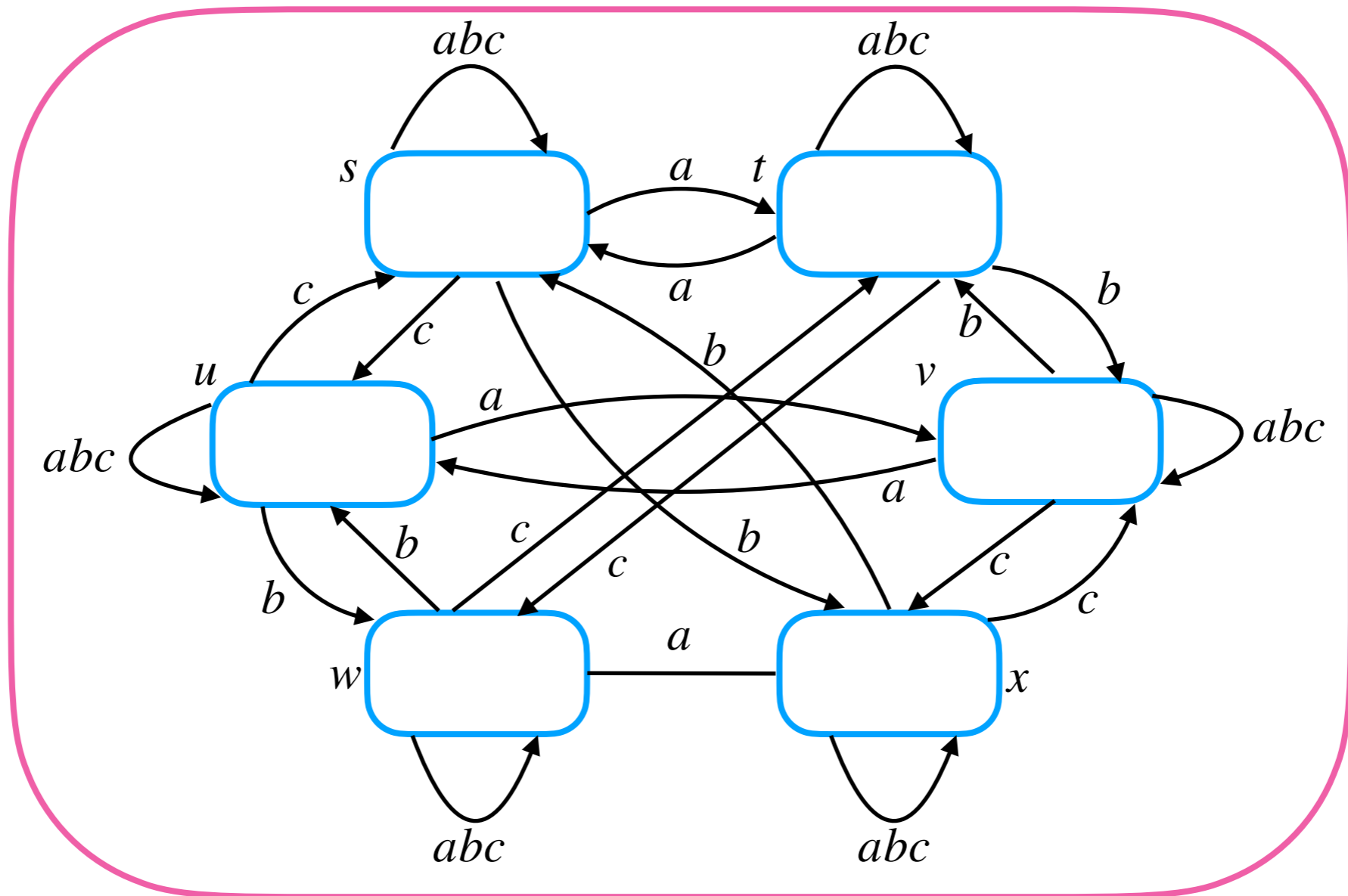
An **equivalence relation** is a binary relation that is **reflexive**, **symmetric** and **transitive**.



A Quick Aside

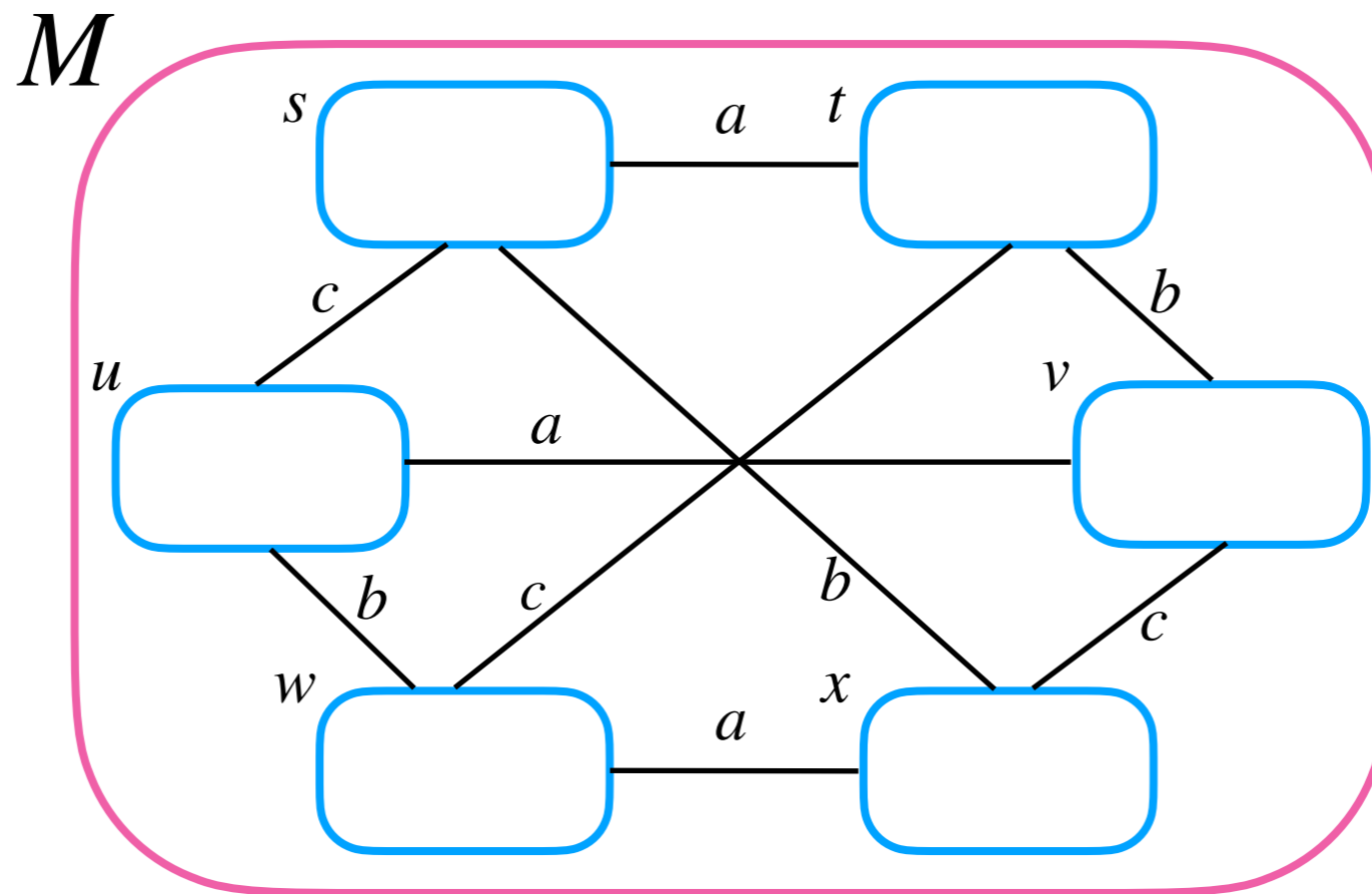
An **equivalence relation** is a binary relation that is **reflexive**, **symmetric** and **transitive**.

M



A Quick Aside

An **equivalence relation** is a binary relation that is **reflexive**, **symmetric** and **transitive**.

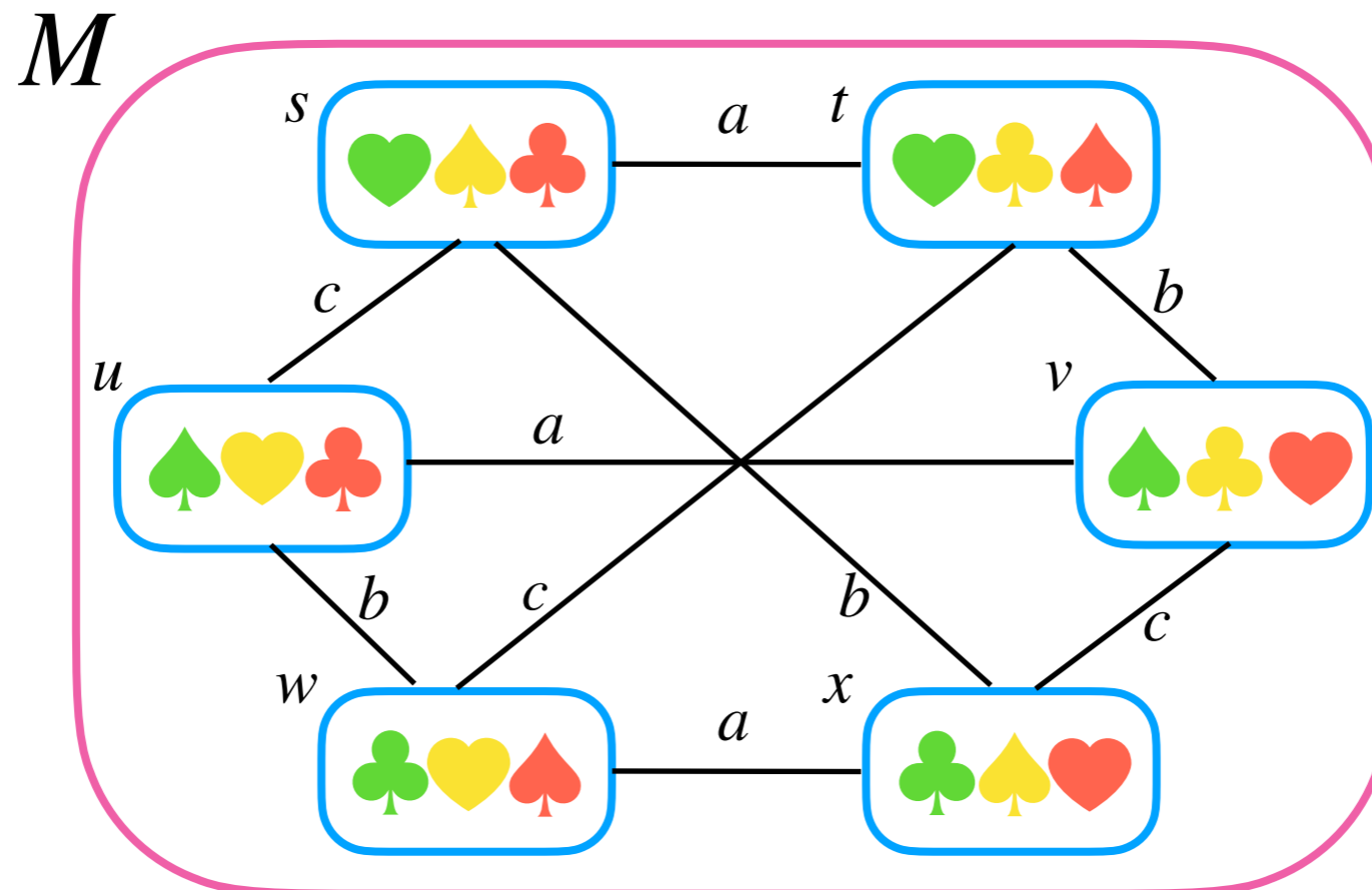


Epistemic Logic

Epistemic models

An **epistemic model** M is a tuple (S, \sim, V) , where

- $S \neq \emptyset$ is a set of states;
- $\sim: A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an equivalence relation;
- $V: P \rightarrow 2^S$ is the valuation function.



Epistemic Logic

From Greek **episteme** that means **knowledge**

Language of EL $\mathcal{EL} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi$

Epistemic models

An **epistemic model** M is a tuple (S, \sim, V) , where

- $S \neq \emptyset$ is a set of states;
- $\sim: A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an equivalence relation;
- $V: P \rightarrow 2^S$ is the valuation function.

Pointed model

A pair of M and $s \in S$ is called a **pointed model** and is denoted as M_s

Semantics of EL

$$M_s \models p \text{ iff } s \in V(p)$$

$$M_s \models \neg\varphi \text{ iff } M_s \not\models \varphi$$

$$M_s \models \varphi \wedge \psi \text{ iff } M_s \models \varphi \text{ and } M_s \models \psi$$

$$M_s \models \Box_a \varphi \text{ iff } \forall t \in S : s \sim_a t \text{ implies } M_t \models \varphi$$

$$M_s \models \Diamond_a \varphi \text{ iff } \exists t \in S : s \sim_a t \text{ and } M_t \models \varphi$$

Note that $\Diamond_a \varphi$ is equivalent to $\neg \Box_a \neg \varphi$

$\psi \vee \varphi$ is equivalent to $\neg(\neg\psi \wedge \neg\varphi)$

$\psi \rightarrow \varphi$ is equivalent to $\neg\psi \vee \varphi$

Properties of Knowledge

I. What is known is true

$\Box_a \varphi \rightarrow \varphi$ is **valid** (is a law of EL)

Corresponds to **reflexivity**

What do you think about belief?

Properties of Knowledge

I. What is known is true

II. Positive introspection

If I know φ , then I know that I know φ

$\Box_a \varphi \rightarrow \Box_a \Box_a \varphi$ is **valid** (is a law of EL)

Corresponds to **transitivity**

Properties of Knowledge

I. What is known is true

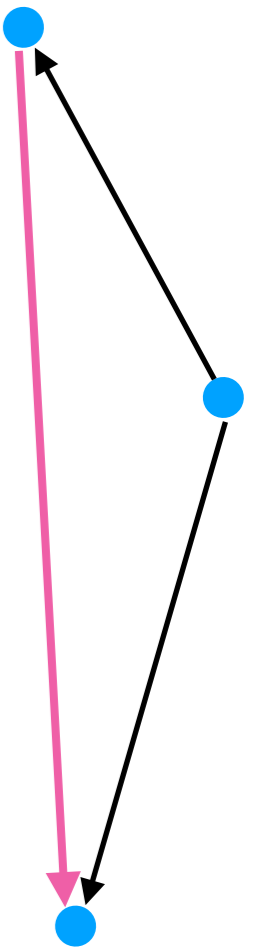
II. Positive introspection

III. Negative introspection

If I don't know φ , then I know that I don't know φ

$\neg \Box_a \varphi \rightarrow \Box_a \neg \Box_a \varphi$ is **valid**

Corresponds to **euclidity**



Properties of Knowledge

I. What is known is true

$$\Box_a \varphi \rightarrow \varphi$$

II. Positive introspection

$$\Box_a \varphi \rightarrow \Box_a \Box_a \varphi$$

III. Negative introspection

$$\neg \Box_a \varphi \rightarrow \Box_a \neg \Box_a \varphi$$

Theorem. I, II, and III are true everywhere in a model iff agents' relations in that model are equivalences

Truth of logical laws I, II, and III

Equivalence condition on models

Axiomatisation of EL

Propositional tautologies

$$\Box_a (\varphi \rightarrow \psi) \rightarrow (\Box_a \varphi \rightarrow \Box_a \psi)$$

$$\Box_a \varphi \rightarrow \varphi \quad \text{Reflexivity}$$

$$\Box_a \varphi \rightarrow \Box_a \Box_a \varphi \quad \text{Transitivity}$$

$$\neg \Box_a \varphi \rightarrow \Box_a \neg \Box_a \varphi \quad \text{Euclid}$$

From $\varphi, \varphi \rightarrow \psi$ infer ψ

From φ infer $\Box_a \varphi$

Theorem. EL is sound and complete

Theorem. Complexity of SAT-EL is PSPACE-complete

Satisfiability: for a given φ , determine whether there is a M_s such that $M_s \models \varphi$

Axiomatisation of EL

Propositional tautologies

$$\Box_a (\varphi \rightarrow \psi) \rightarrow (\Box_a \varphi \rightarrow \Box_a \psi)$$

$$\Box_a \varphi \rightarrow \varphi \quad \text{Reflexivity}$$

$$\Box_a \varphi \rightarrow \Box_a \Box_a \varphi \quad \text{Transitivity}$$

$$\neg \Box_a \varphi \rightarrow \Box_a \neg \Box_a \varphi \quad \text{Euclid}$$

From $\varphi, \varphi \rightarrow \psi$ infer ψ

From φ infer $\Box_a \varphi$

Theorem. EL is sound and complete

Theorem. Complexity of SAT-EL is PSPACE-complete

Theorem. Complexity of MC-EL is P-complete

Model checking: for a given φ and M_s , determine whether $M_s \models \varphi$

Overview of EL

- Extends propositional logic with constructs $\Box_a \varphi$ that mean 'agent a knows φ '
- Interpreted on (epistemic) models that consist of states, equivalence relations for each agent, and truth assignment of atomic propositions
- Knowledge is assumed to be truthful, and obey positive and negative introspections
- EL allows one to reason not only about knowledge of simple facts, but about higher-order knowledge as well

Further research in EL

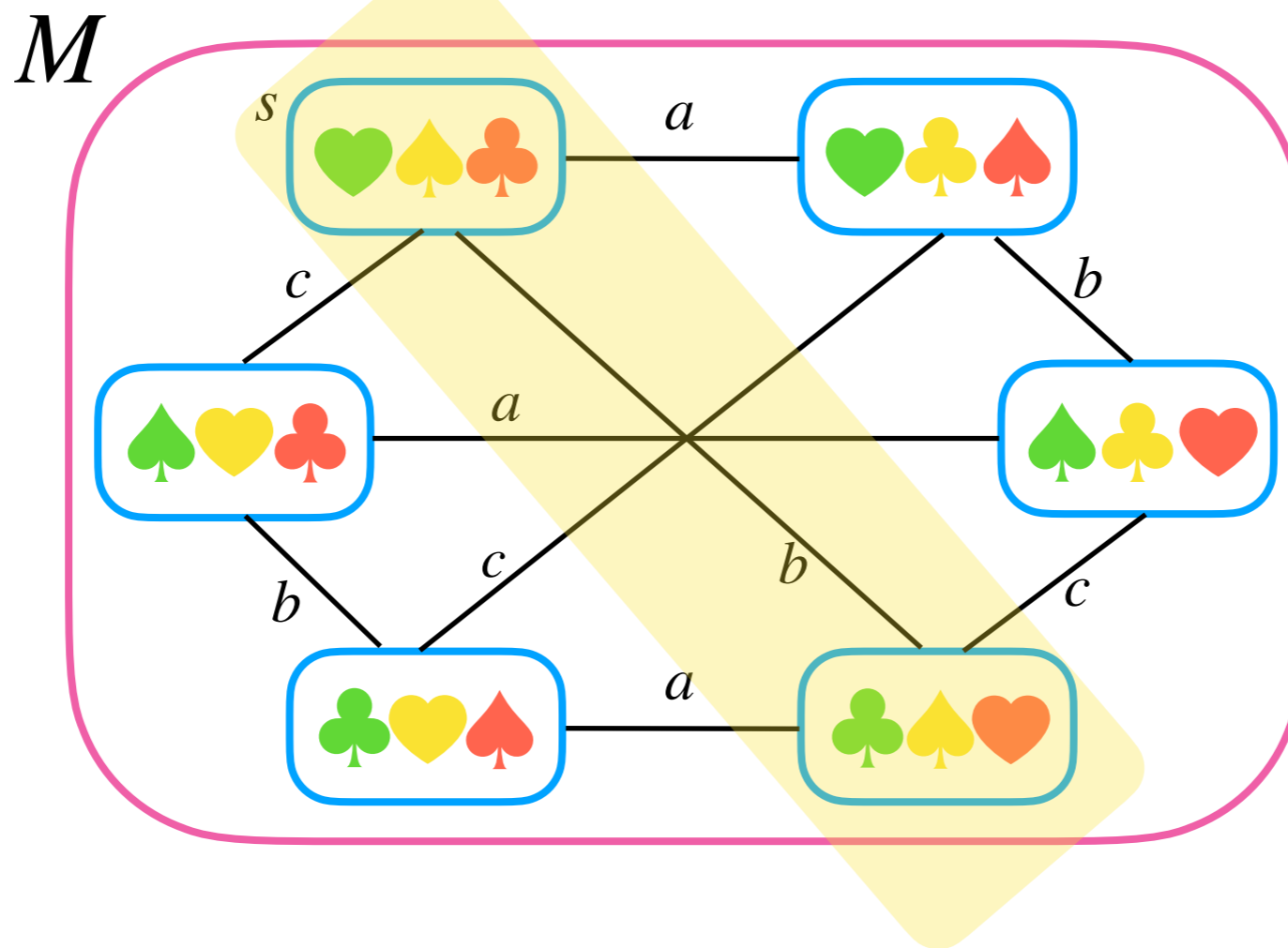
- More appropriate notions of **knowledge and belief**
- Knowledge and belief of **groups of agents**
- Applications to **epistemic game theory**
- Epistemic analysis of **CS protocols**, e.g. gossip protocol and dining cryptographers
- **AI agents**, e.g. BDI architecture and epistemic planning
- And so on and so on and so on and so on...

Part II

Public Announcement Logic

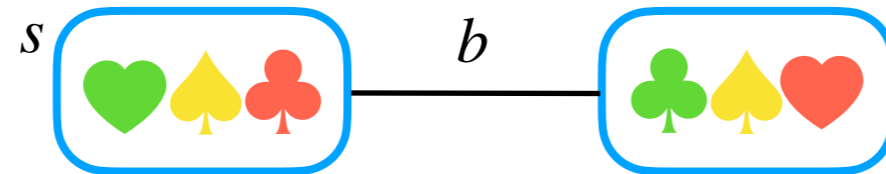
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs

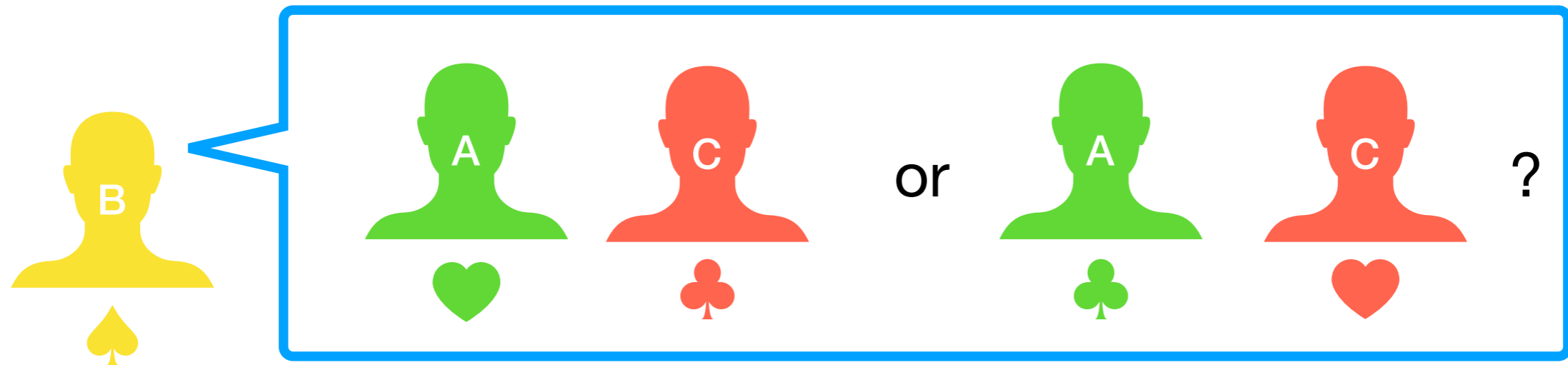


Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs

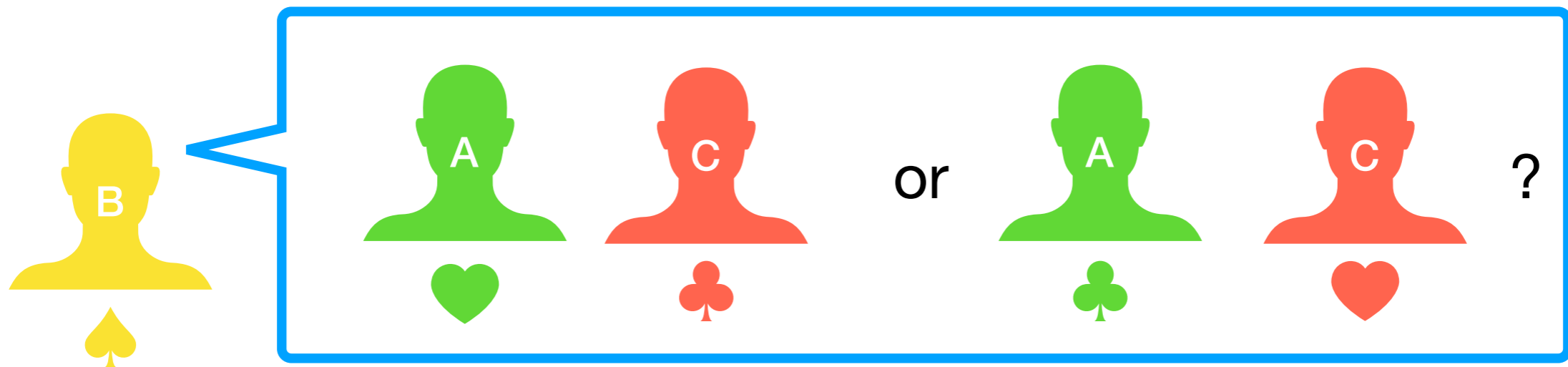
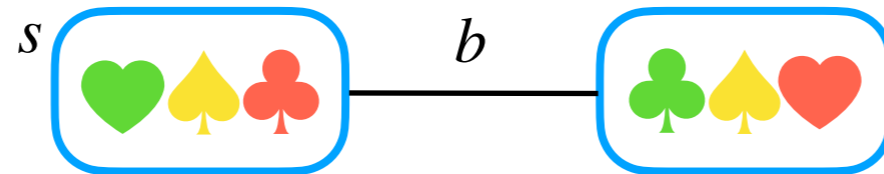


Initial situation



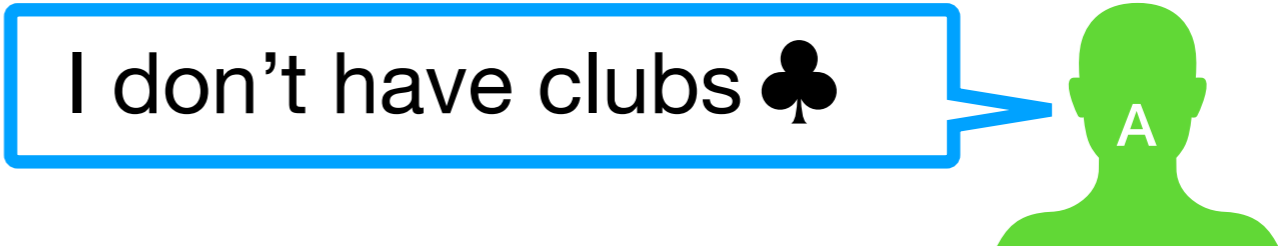
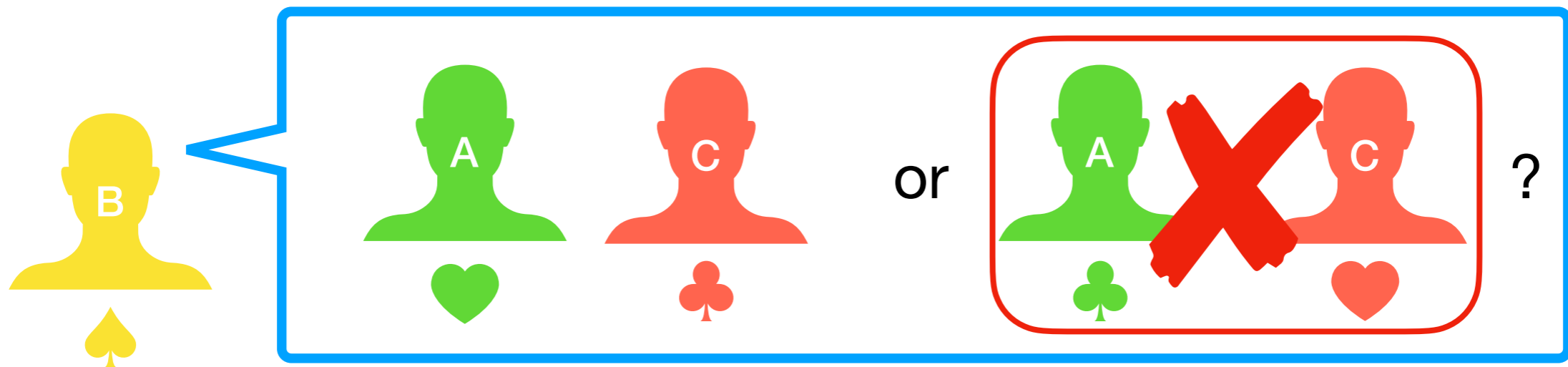
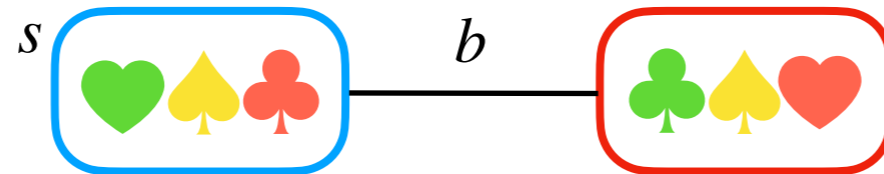
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



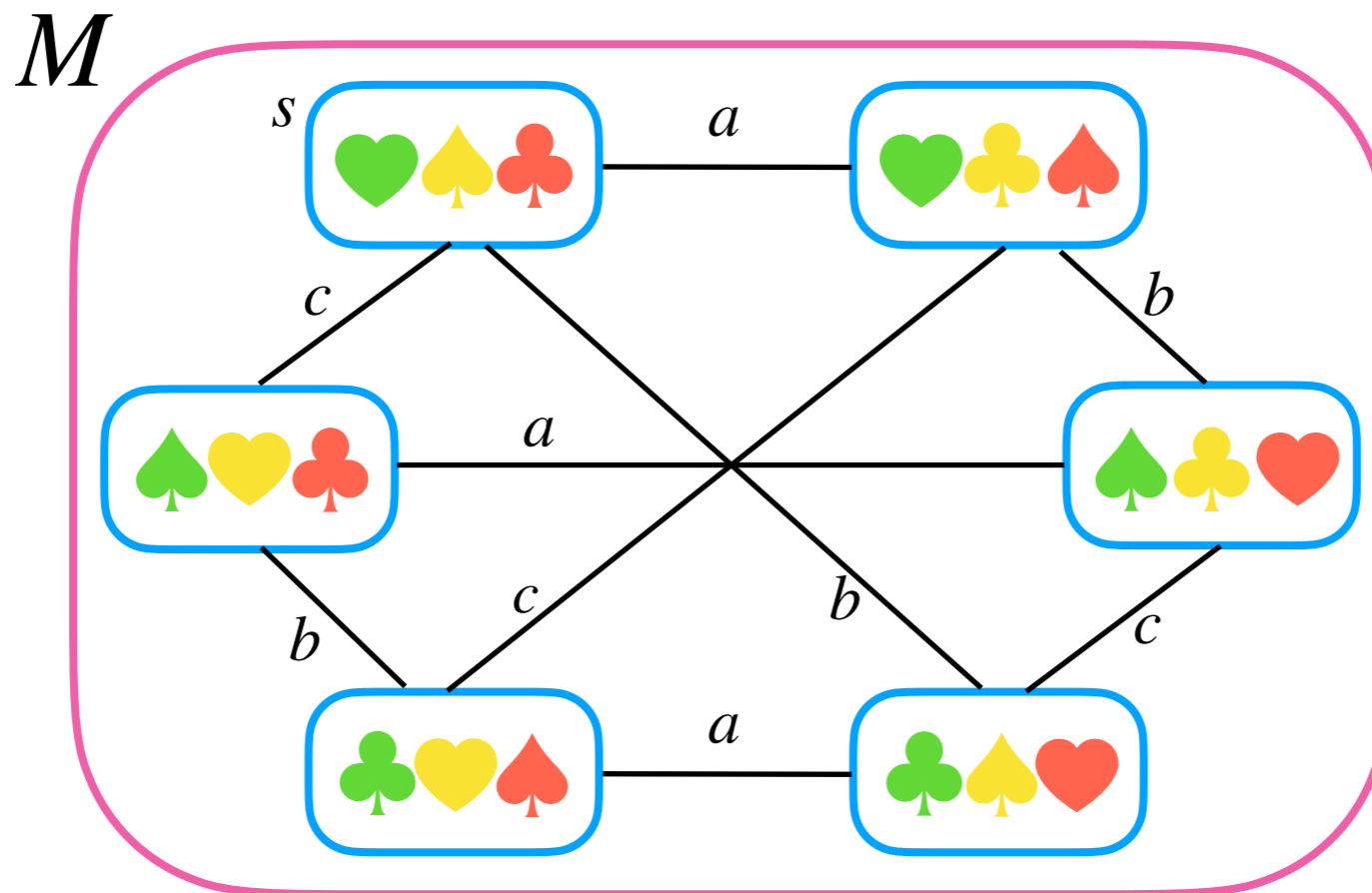
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Card Example

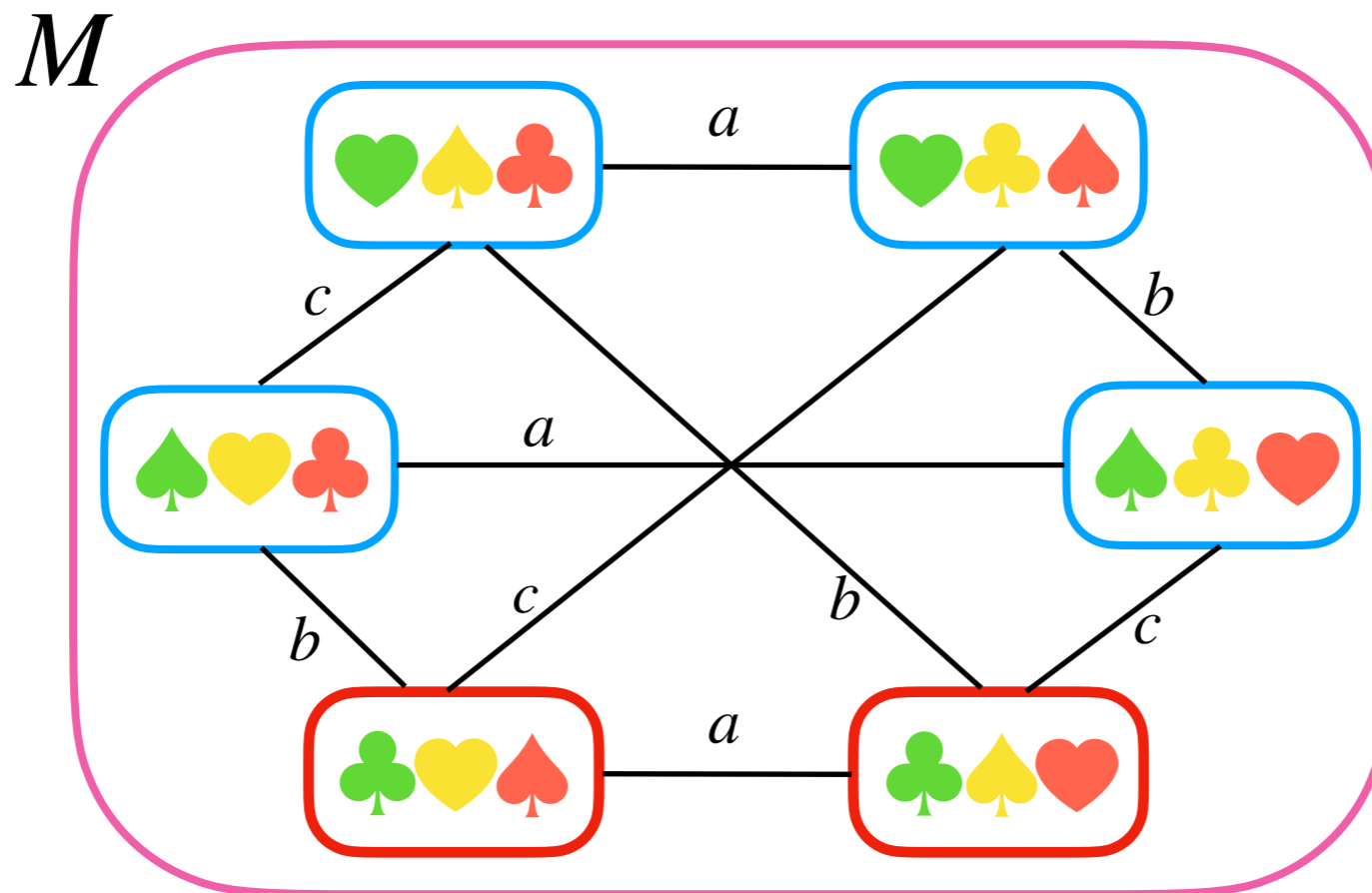
Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Alice says that she does not have clubs: $\neg \clubsuit_a$

Card Example

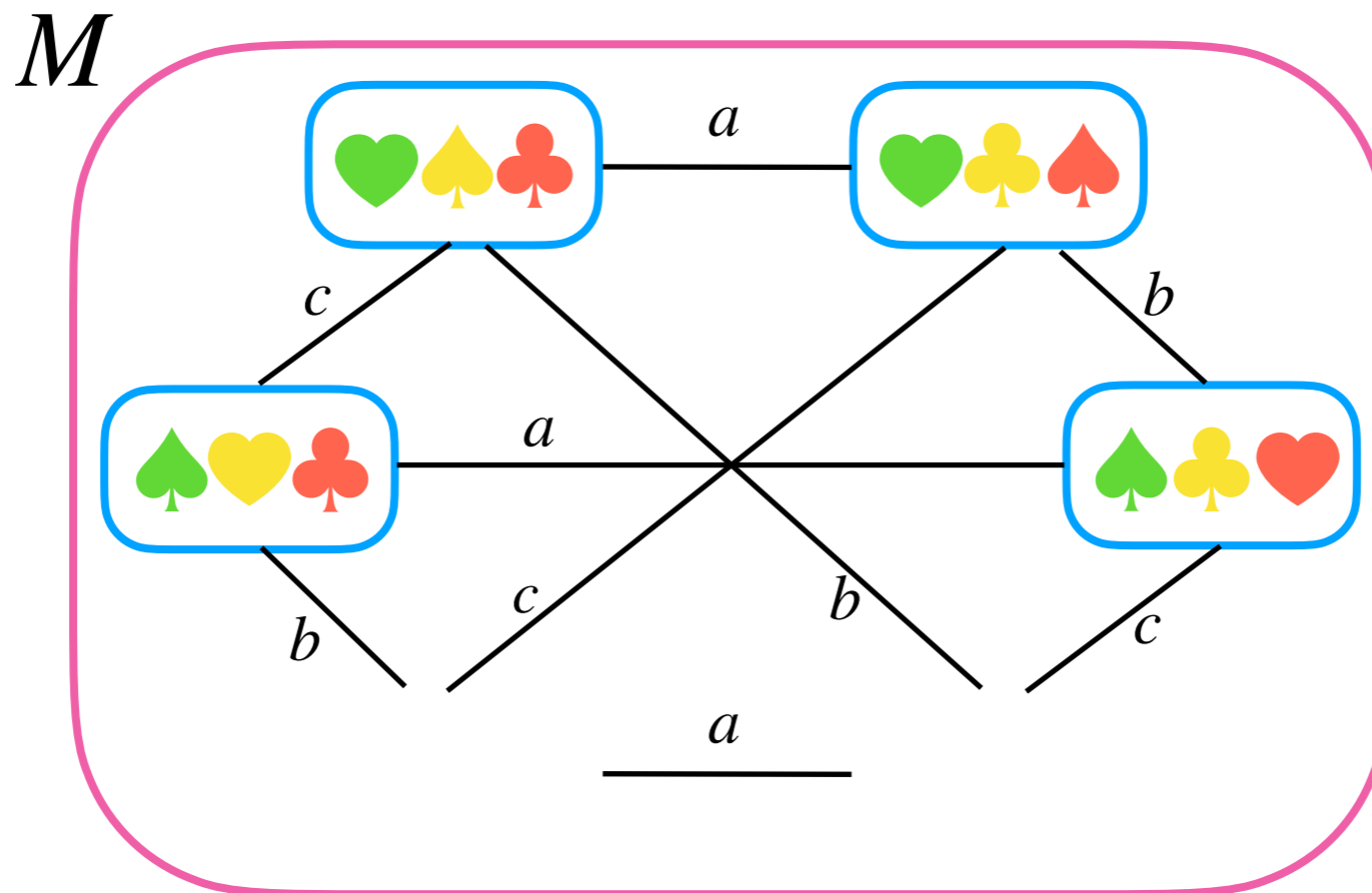
Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Alice says that she does not have clubs: $\neg \clubsuit_a$

Card Example

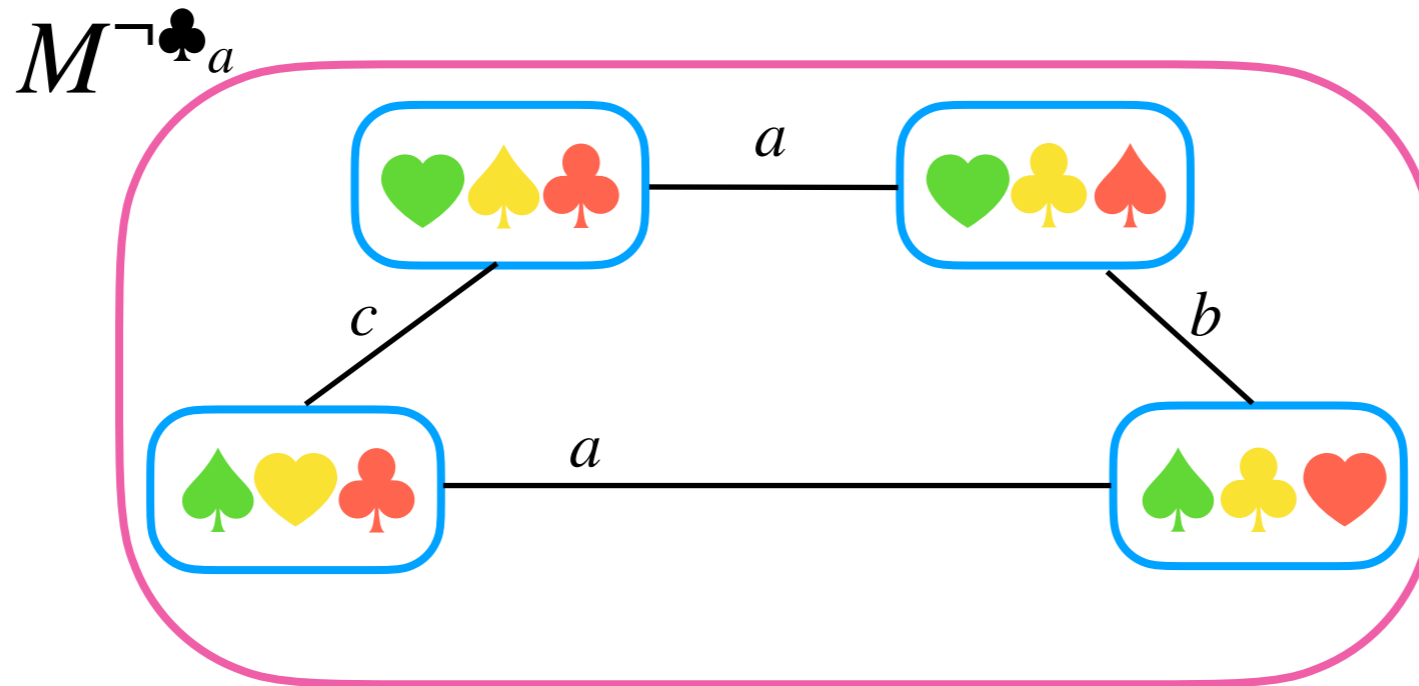
Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Alice says that she does not have clubs: $\neg \clubsuit_a$

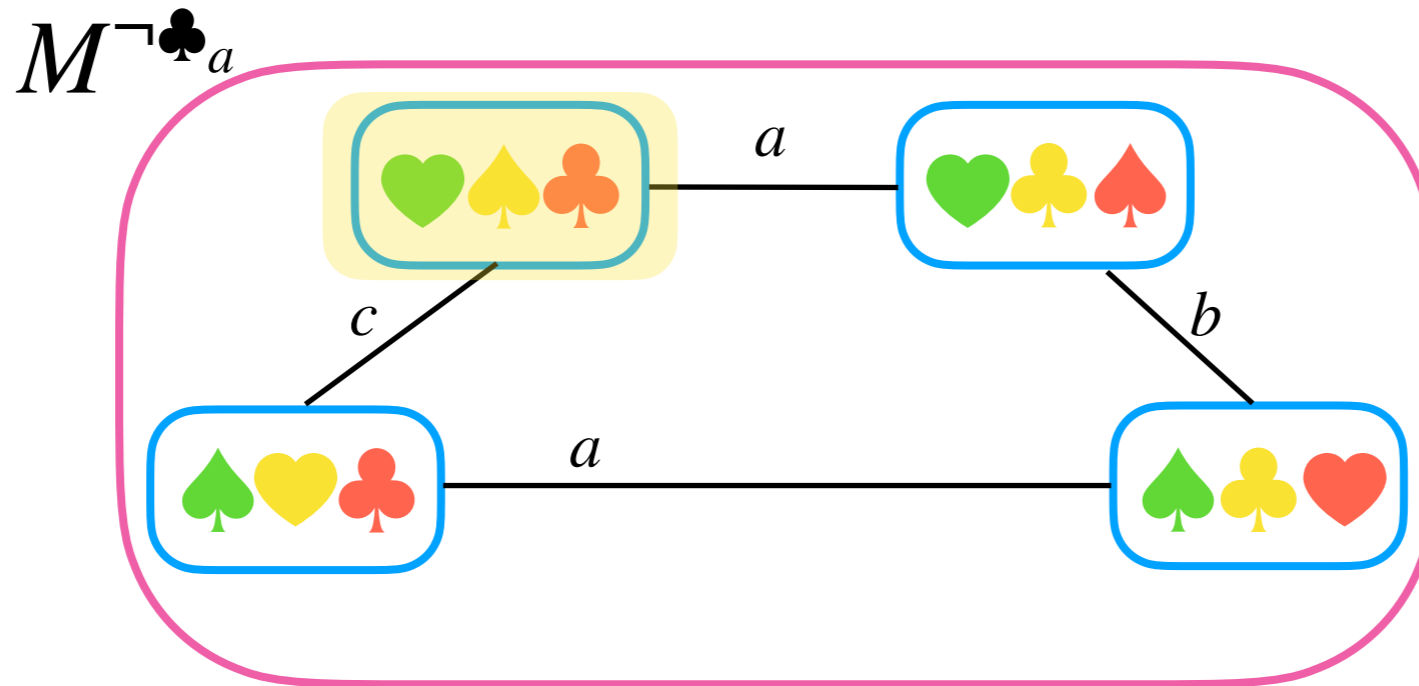
Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Card Example

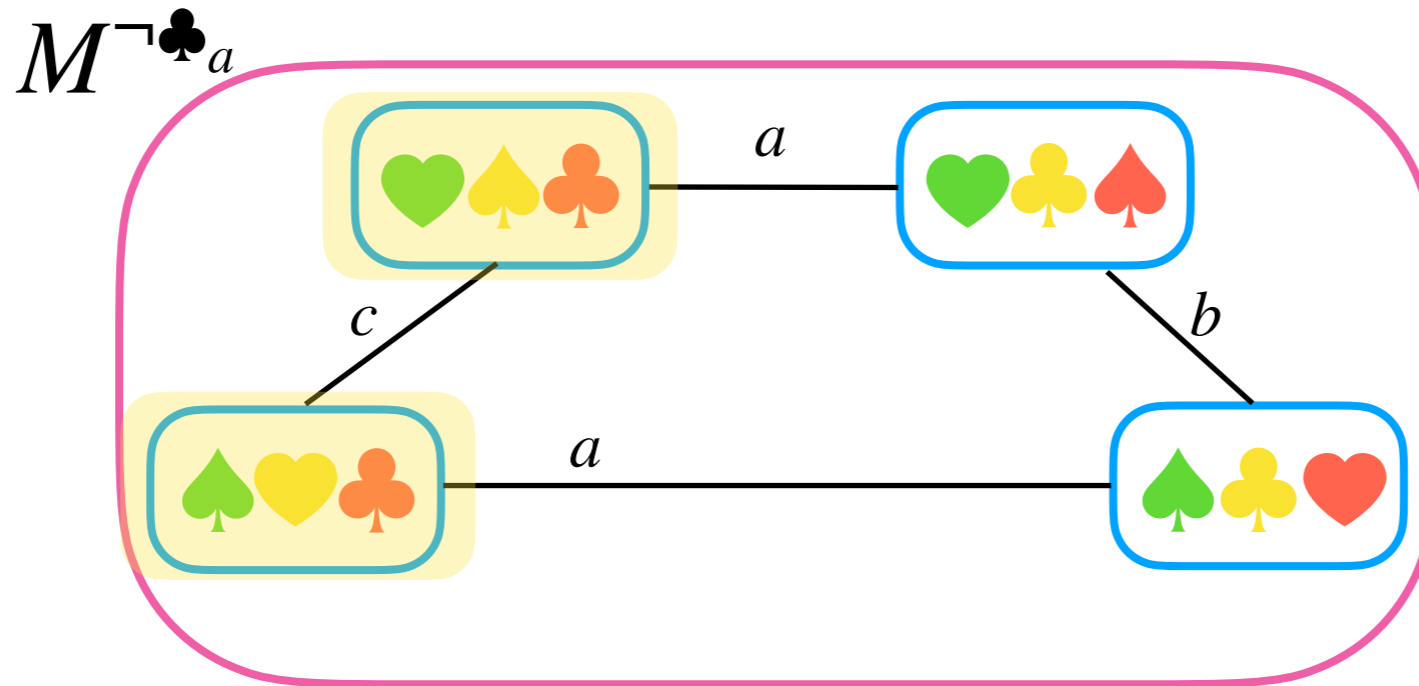
Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Bob says that he now knows that **Carol** has clubs: $\Box_b \clubsuit_c$

Card Example

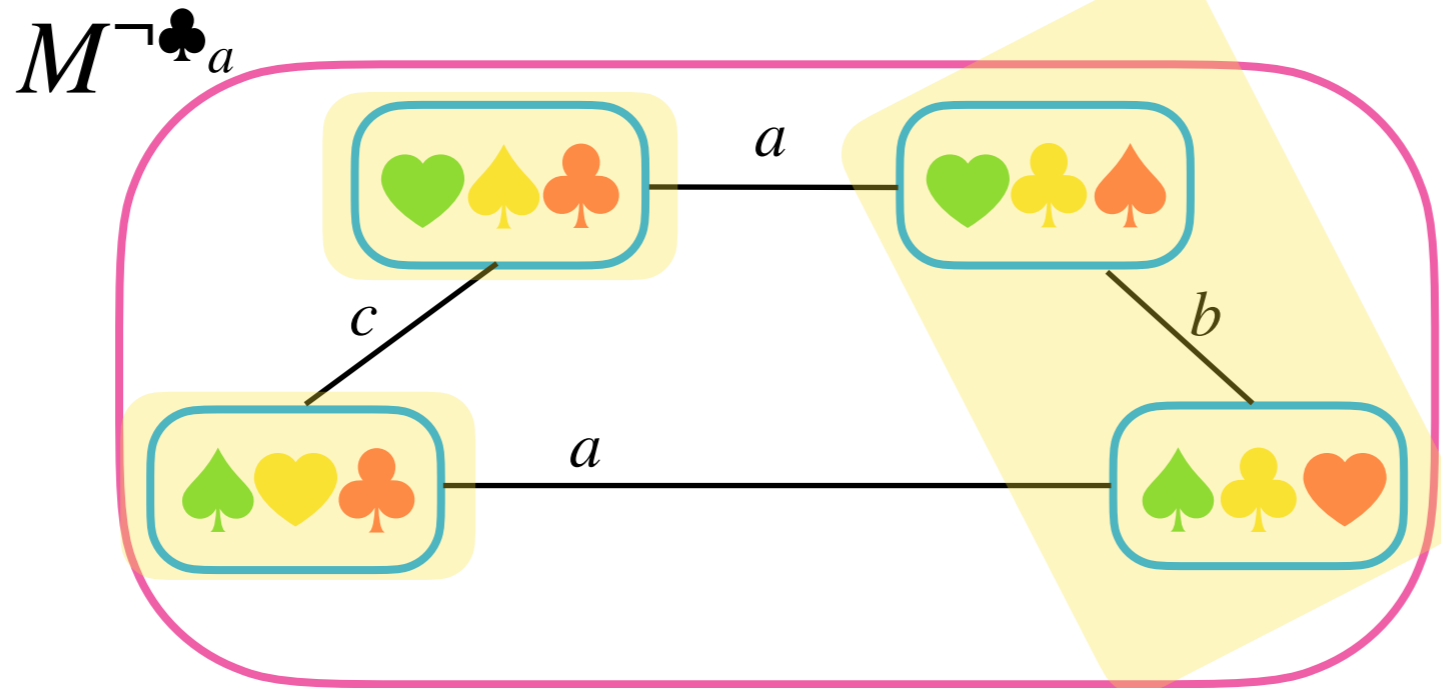
Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Bob says that he now knows that **Carol** has clubs: $\Box_b \clubsuit_c$

Card Example

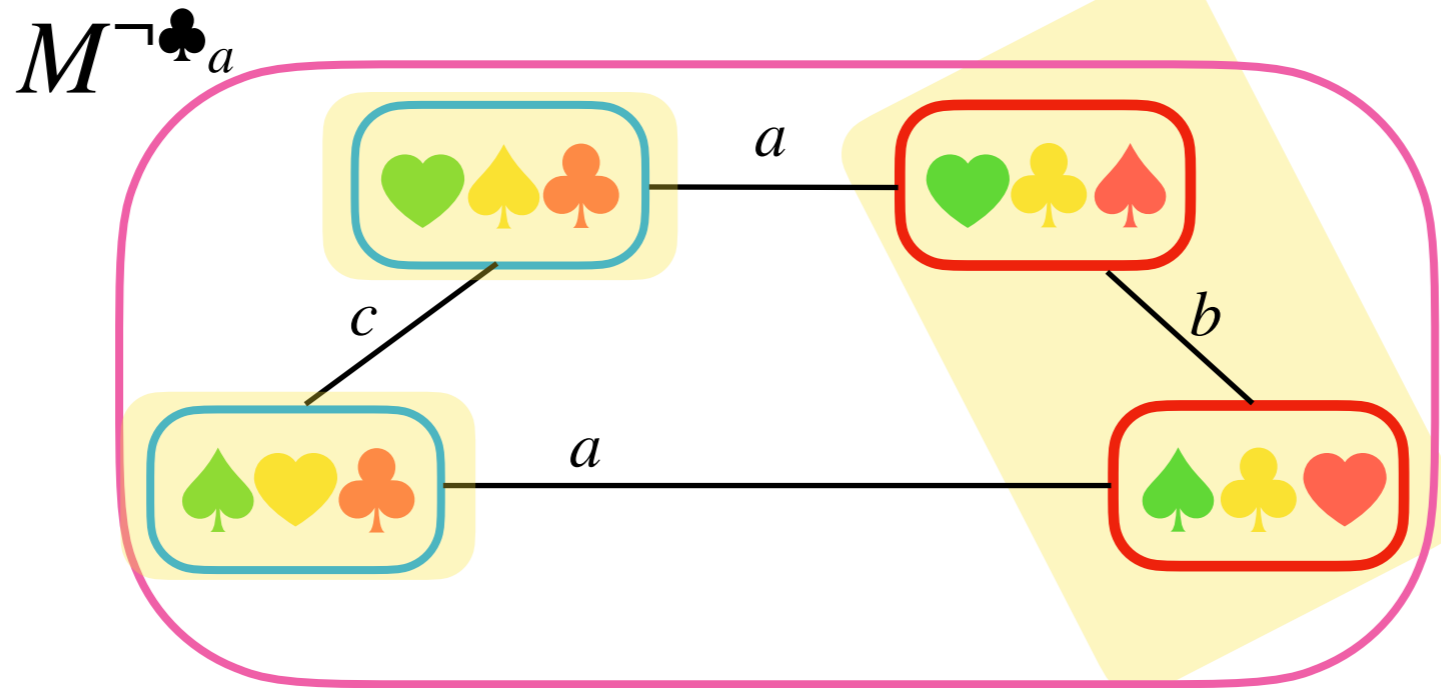
Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



Bob says that he now knows that **Carol** has clubs: $\Box_b \clubsuit_c$

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs

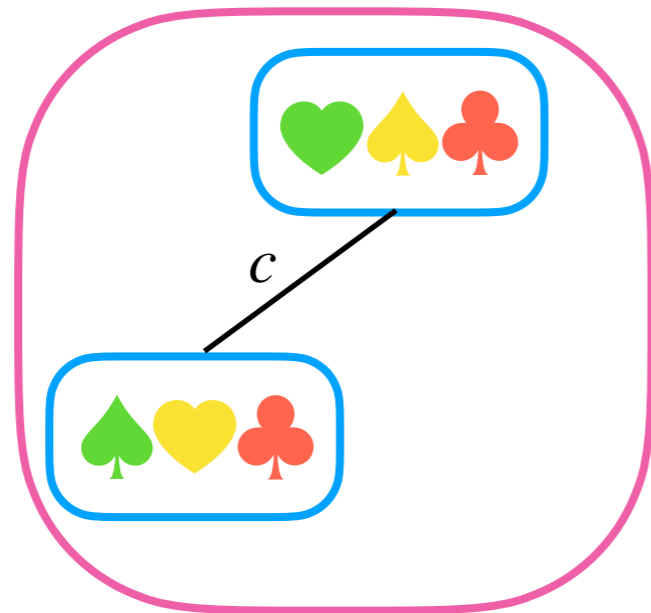


Bob says that he now knows that **Carol** has clubs: $\Box_b \clubsuit_c$

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs

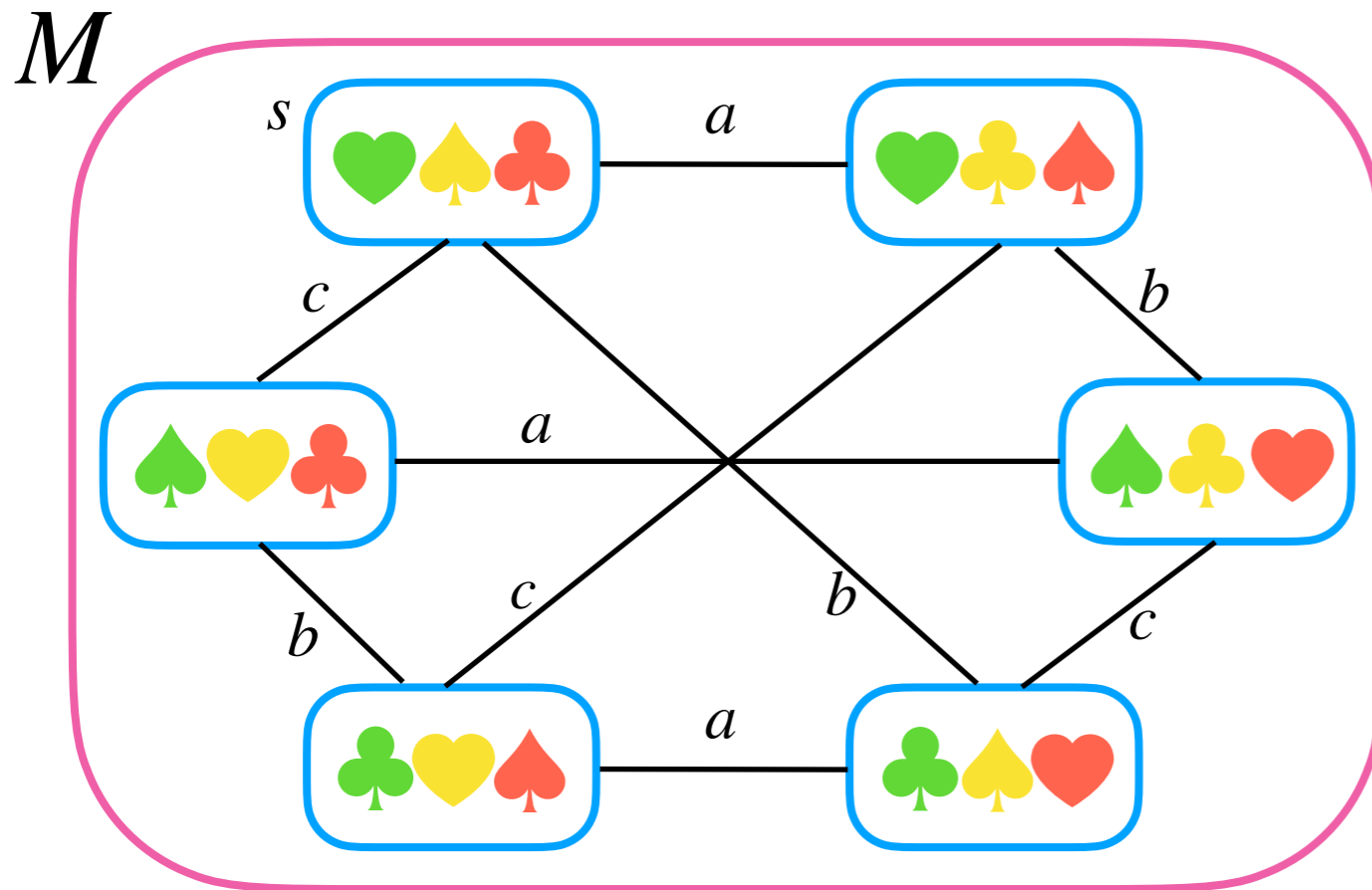
$$M \neg \clubsuit_a, \square_b \clubsuit_c$$



Bob says that he now knows that **Carol** has clubs: $\square_b \clubsuit_c$

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



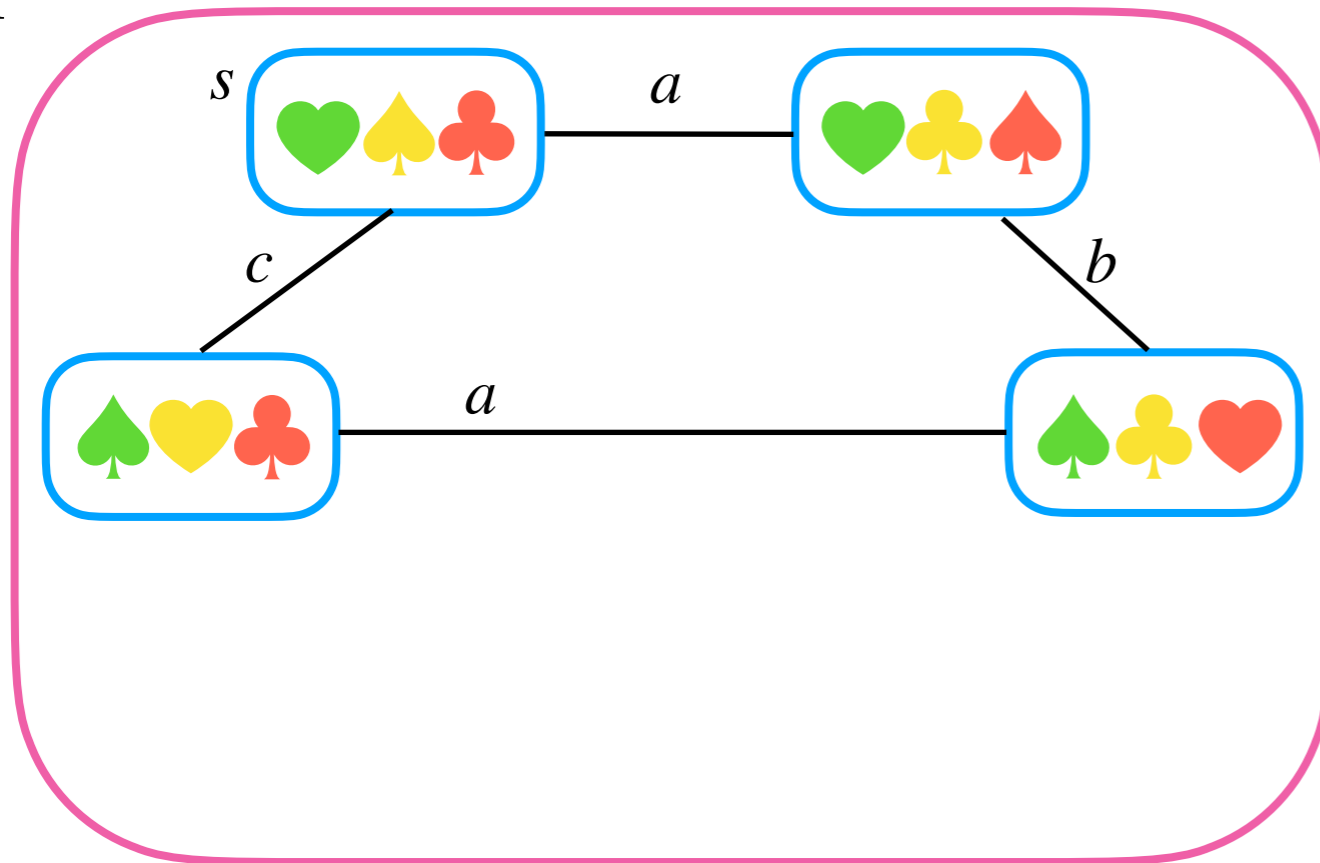
$$M_s \models [\neg \clubsuit_a] \Box_b (\heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c)$$

$[\psi]\varphi$: after **public announcement** of ψ , φ is true

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs

$M^{\neg\clubsuit_a}$



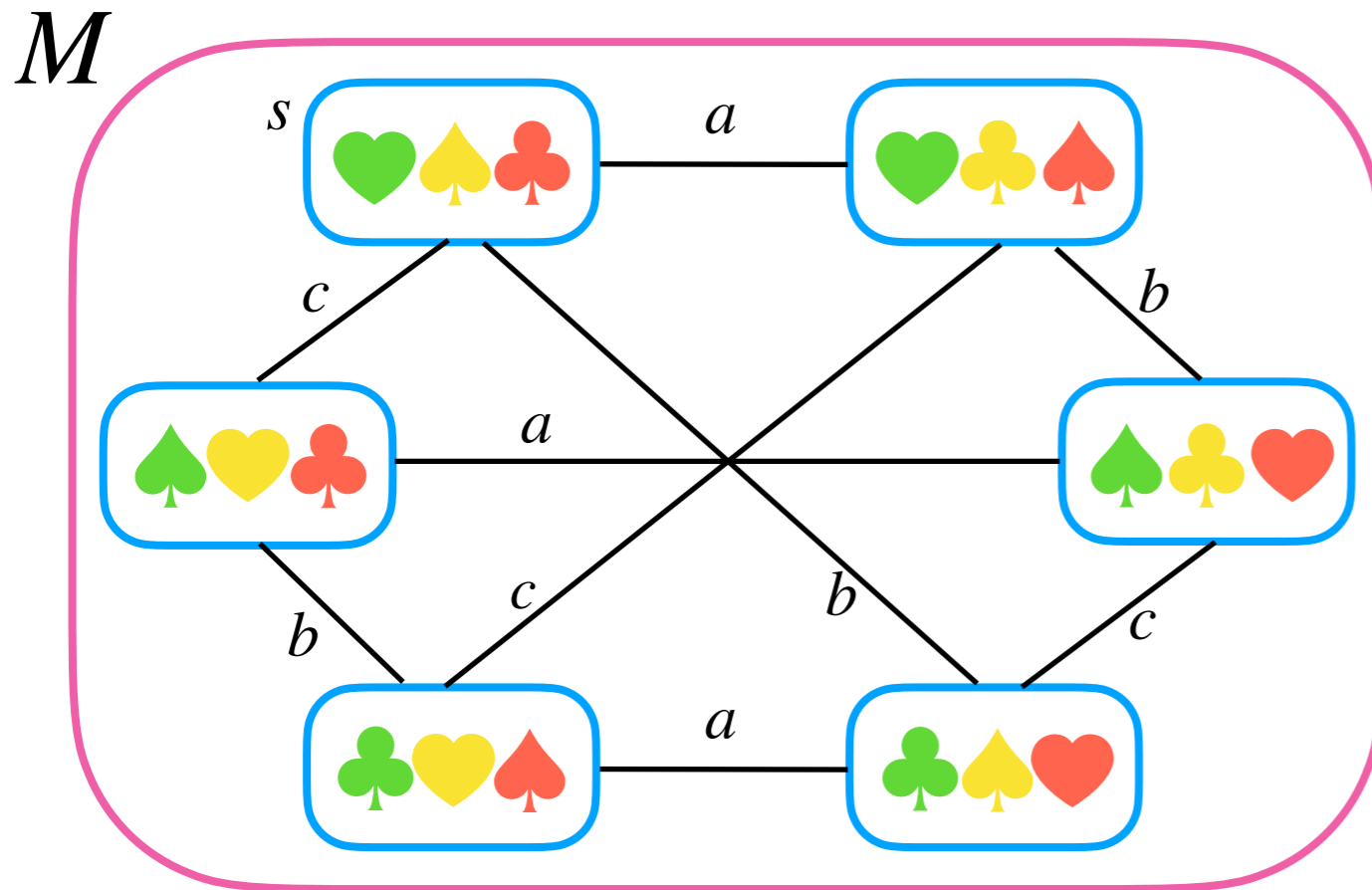
$$M_s \models [\neg\clubsuit_a] \Box_b (\heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c)$$

$$M_s^{\neg\clubsuit_a} \models \Box_b (\heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c)$$

$[\psi]\varphi$: after **public announcement** of ψ , φ is true

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



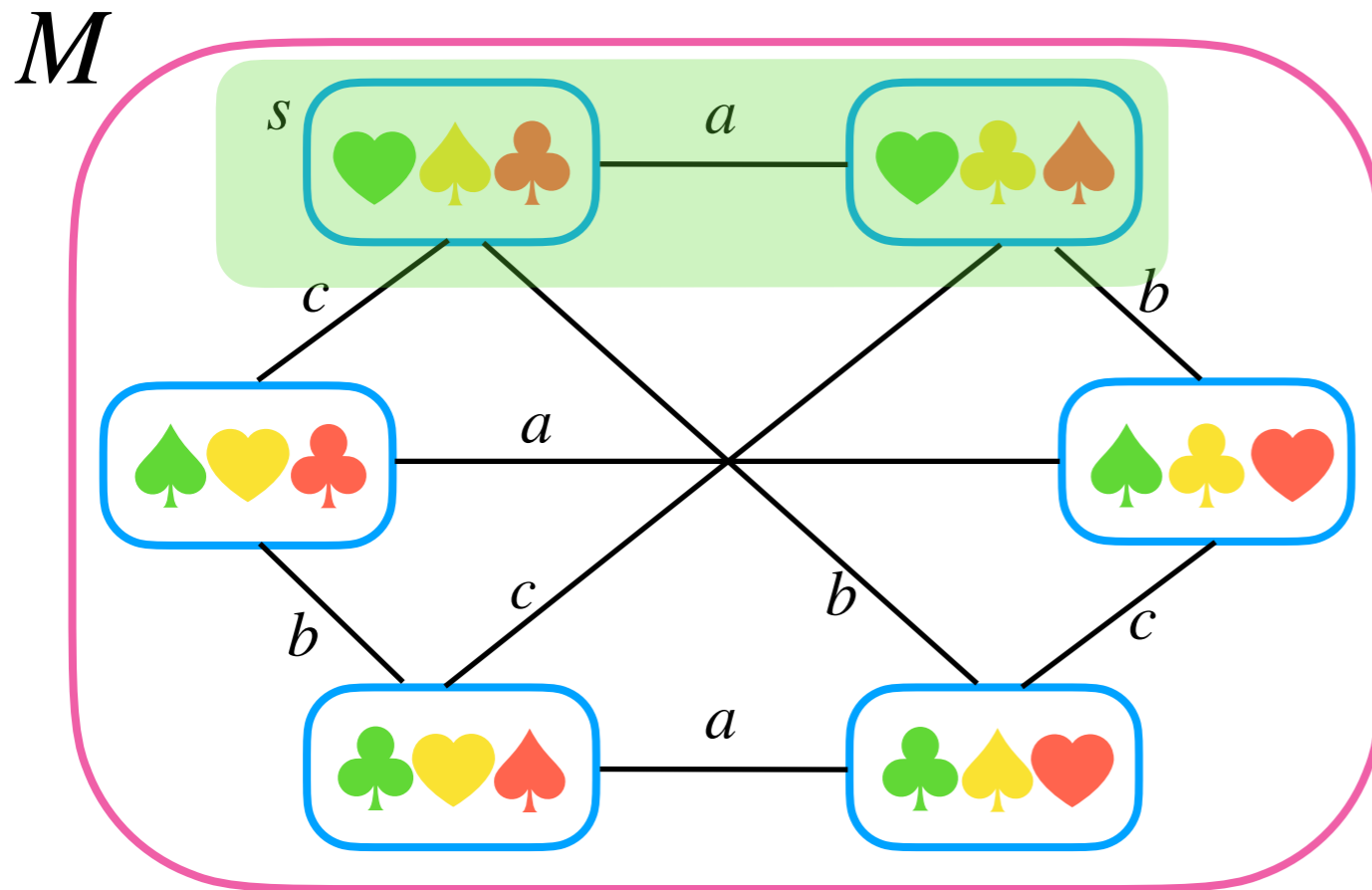
$$M_s \models [\neg \clubsuit_a] \Box_b (\heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models [\Box_a \neg \heartsuit_c] \Box_c \heartsuit_a$$

$[\psi]\varphi$: after **public announcement** of ψ , φ is true

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



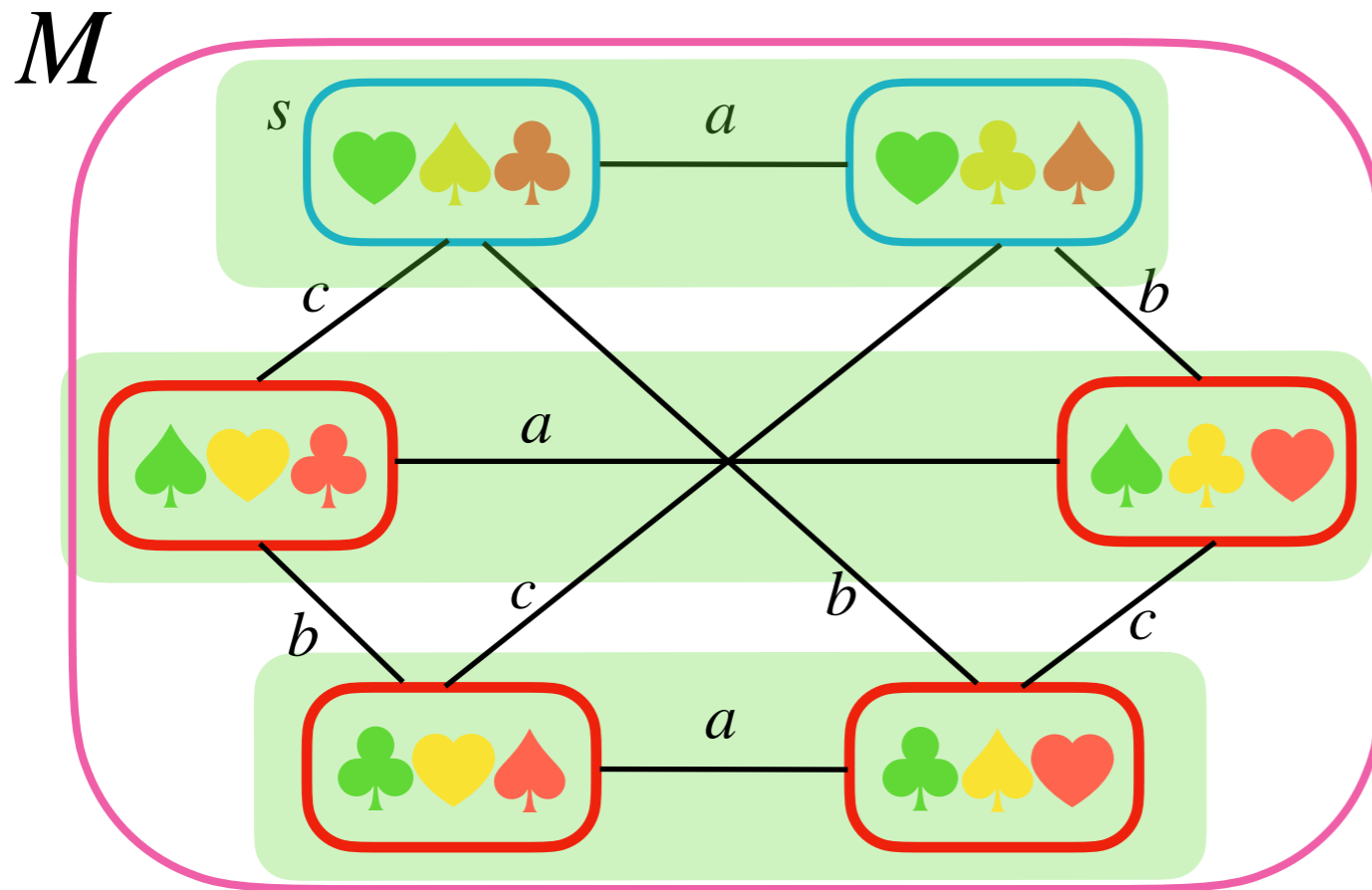
$$M_s \models [\neg \clubsuit_a] \Box_b (\heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models [\Box_a \neg \heartsuit_c] \Box_c \heartsuit_a$$

$[\psi]\varphi$: after **public announcement** of ψ , φ is true

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs



$$M_s \models [\neg \clubsuit_a] \Box_b (\heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c)$$

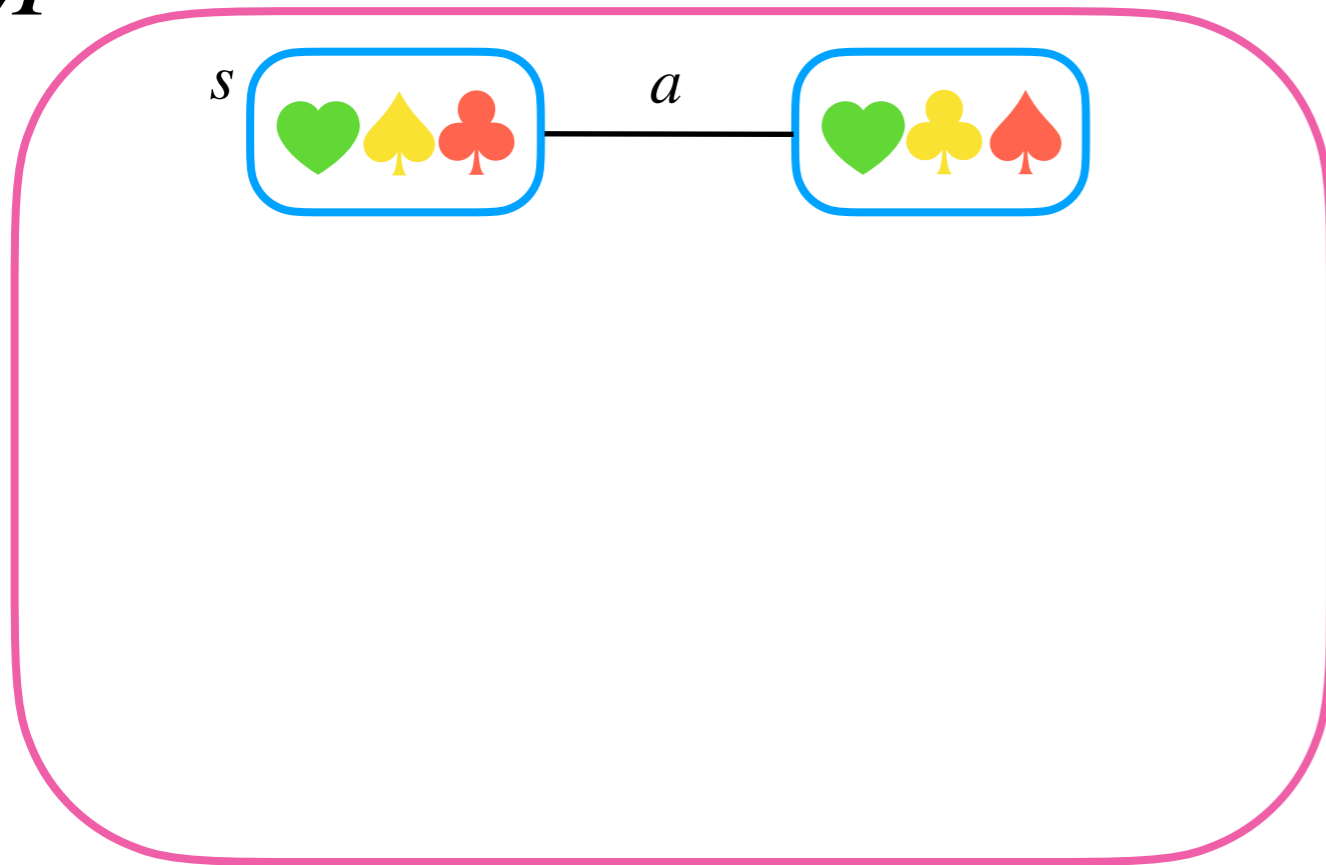
$$M_s \models [\Box_a \neg \heartsuit_c] \Box_c \heartsuit_a$$

$[\psi]\varphi$: after **public announcement** of ψ , φ is true

Card Example

Three agents, **Alice**, **Bob**, and **Carol**, have each drawn one card from a deck of {♥ ♠ ♣}, and then **Alice** says that she does not have clubs

$$M \square_a \neg \heartsuit_c$$



$$M_s \models [\neg \clubsuit_a] \square_b (\heartsuit_a \wedge \spadesuit_b \wedge \clubsuit_c)$$

$$M_s \models [\square_a \neg \heartsuit_c] \square_c \heartsuit_a$$

$[\psi]\varphi$: after **public announcement** of ψ , φ is true

Public Announcement Logic

Language of
PAL

$$\mathcal{PAL} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi \mid [\varphi]\varphi$$

Semantics

$$M_s \vDash [\psi]\varphi \text{ iff } M_s \vDash \psi \text{ implies } M_s^\psi \vDash \varphi$$

$$M_s \vDash \langle \psi \rangle \varphi \text{ iff } M_s \vDash \psi \text{ and } M_s^\psi \vDash \varphi$$

Updated model

Let $M = (S, \sim, V)$ and $\varphi \in \mathcal{PAL}$. An **updated model** M^φ is a tuple $(S^\varphi, \sim^\varphi, V^\varphi)$, where

- $S^\varphi = \{s \in S \mid M_s \vDash \varphi\}$;
- $\sim_a^\varphi = \sim_a \cap (S^\varphi \times S^\varphi)$;
- $V^\varphi(p) = V(p) \cap S^\varphi$.

Overview of PAL So Far

- **Public announcement** is an event of all agents publicly and simultaneously learning some true piece of information
- Public announcements are not necessarily speech acts, they can be acts of publishing, posting, sharing, etc.
- **Fun fact:** public announcements do not necessarily remain true after being announced. ‘My birthday is in November, and you don’t know this’
- How much **expressivity** do they add, compared to the standard EL?

Overview of PAL So Far

- **Public announcement** is an event of all agents publicly and simultaneously learning some true piece of information
- Public announcements are not necessarily speech acts, they can be acts of publishing, posting, sharing, etc.
- **Fun fact:** public announcements do not necessarily remain true after being announced. ‘My birthday is in November, and you don’t know this’
- How much **expressivity** do they add, compared to the standard EL? **None at all!**

Properties of Public Announcements

Consider the validities (laws) of PAL

$$[\varphi]p \leftrightarrow (\varphi \rightarrow p)$$

$$[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$$

$$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$$

$$[\varphi]\Box_a\psi \leftrightarrow (\varphi \rightarrow \Box_a[\varphi]\psi)$$

$$[\varphi][\psi]\chi \leftrightarrow ([\varphi \wedge [\varphi]\psi]\chi)$$

These rewriting rules decrease the complexity of a formula

Example

$$[\Box_a p] \neg \Box_b q$$

Properties of Public Announcements

Consider the validities (laws) of PAL

$$[\varphi]p \leftrightarrow (\varphi \rightarrow p)$$

$$[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$$

$$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$$

$$[\varphi]\Box_a\psi \leftrightarrow (\varphi \rightarrow \Box_a[\varphi]\psi)$$

$$[\varphi][\psi]\chi \leftrightarrow ([\varphi \wedge [\varphi]\psi]\chi)$$

These rewriting rules decrease the complexity of a formula

Example

$$[\Box_a p] \neg \Box_b q$$

$$\Box_a p \rightarrow \neg[\Box_a p] \Box_b q$$

Properties of Public Announcements

Consider the validities (laws) of PAL

$$[\varphi]p \leftrightarrow (\varphi \rightarrow p)$$

$$[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$$

$$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$$

$$[\varphi]\Box_a\psi \leftrightarrow (\varphi \rightarrow \Box_a[\varphi]\psi)$$

$$[\varphi][\psi]\chi \leftrightarrow ([\varphi \wedge [\varphi]\psi]\chi)$$

These rewriting rules decrease the complexity of a formula

Example

$$[\Box_a p] \neg \Box_b q$$

$$\Box_a p \rightarrow \neg [\Box_a p] \Box_b q$$

$$\Box_a p \rightarrow \neg (\Box_a p \rightarrow \Box_b [\Box_a p] q)$$

Properties of Public Announcements

Consider the validities (laws) of PAL

$$[\varphi]p \leftrightarrow (\varphi \rightarrow p)$$

$$[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$$

$$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$$

$$[\varphi]\Box_a\psi \leftrightarrow (\varphi \rightarrow \Box_a[\varphi]\psi)$$

$$[\varphi][\psi]\chi \leftrightarrow ([\varphi \wedge [\varphi]\psi]\chi)$$

These rewriting rules decrease the complexity of a formula

Example

$$[\Box_a p] \neg \Box_b q$$

$$\Box_a p \rightarrow \neg [\Box_a p] \Box_b q$$

$$\Box_a p \rightarrow \neg (\Box_a p \rightarrow \Box_b [\Box_a p] q)$$

$$\Box_a p \rightarrow \neg (\Box_a p \rightarrow \Box_b (\Box_a p \rightarrow q))$$

Any **potential worries** with the translation?

Theorem. Any formula with public announcements can be **equivalently rewritten** into a formula without them

Axiomatisation of PAL

Axioms of EL

$$[\varphi]p \leftrightarrow (\varphi \rightarrow p)$$

$$[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$$

$$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$$

$$[\varphi]\Box_a\psi \leftrightarrow (\varphi \rightarrow \Box_a[\varphi]\psi)$$

$$[\varphi][\psi]\chi \leftrightarrow ([\varphi \wedge [\varphi]\psi]\chi)$$

From φ infer $[\psi]\varphi$

Theorem. PAL and EL are equally expressive

Theorem. PAL is sound and complete

Theorem. Complexity of SAT-PAL is PSPACE-complete

Axiomatisation of PAL

Axioms of EL

$$[\varphi]p \leftrightarrow (\varphi \rightarrow p)$$

$$[\varphi]\neg\psi \leftrightarrow (\varphi \rightarrow \neg[\varphi]\psi)$$

$$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$$

$$[\varphi]\Box_a\psi \leftrightarrow (\varphi \rightarrow \Box_a[\varphi]\psi)$$

$$[\varphi][\psi]\chi \leftrightarrow ([\varphi \wedge [\varphi]\psi]\chi)$$

From φ infer $[\psi]\varphi$

Theorem. PAL and EL are equally expressive

Theorem. PAL is sound and complete

Theorem. Complexity of SAT-PAL is PSPACE-complete

Theorem. Complexity of MC-PAL is P-complete

Van Benthem, Kooi. *Reduction axioms for epistemic actions*, 2004.



Lutz. *Complexity and Succinctness of Public Announcement Logic*, 2006.

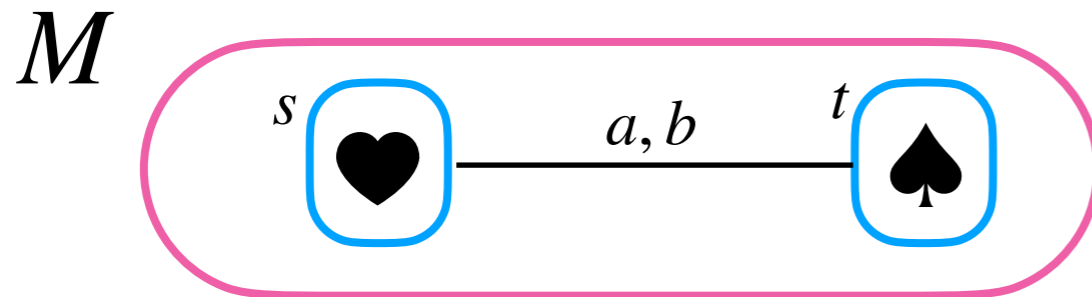
Van Ditmarsch, Van der Hoek, Kooi. *Dynamic Epistemic Logic*, Section 4. 2008.

Part III

Action Models

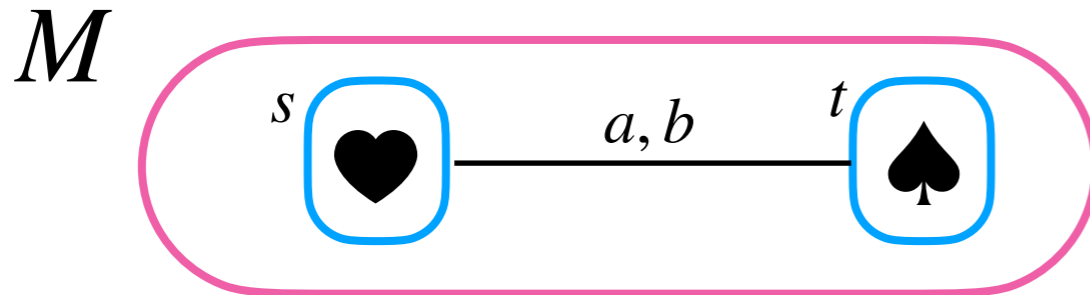
Card Example

There is a card lying face down on a table that can be either  or . Alice and Bob see the card but do not know its suit.



Card Example

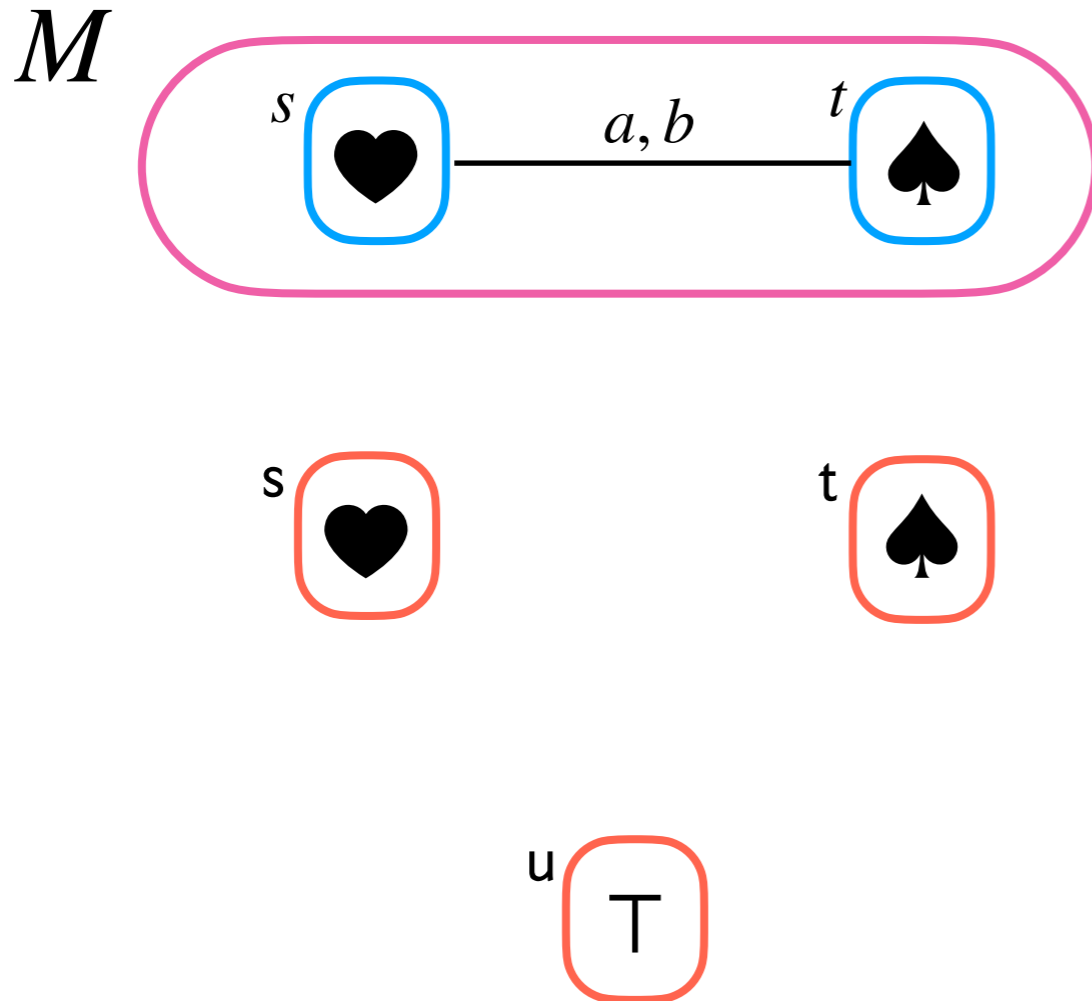
There is a card lying face down on a table that can be either \heartsuit or \spadesuit . Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



Let's take a moment to meditate on 'suspects'...

Card Example

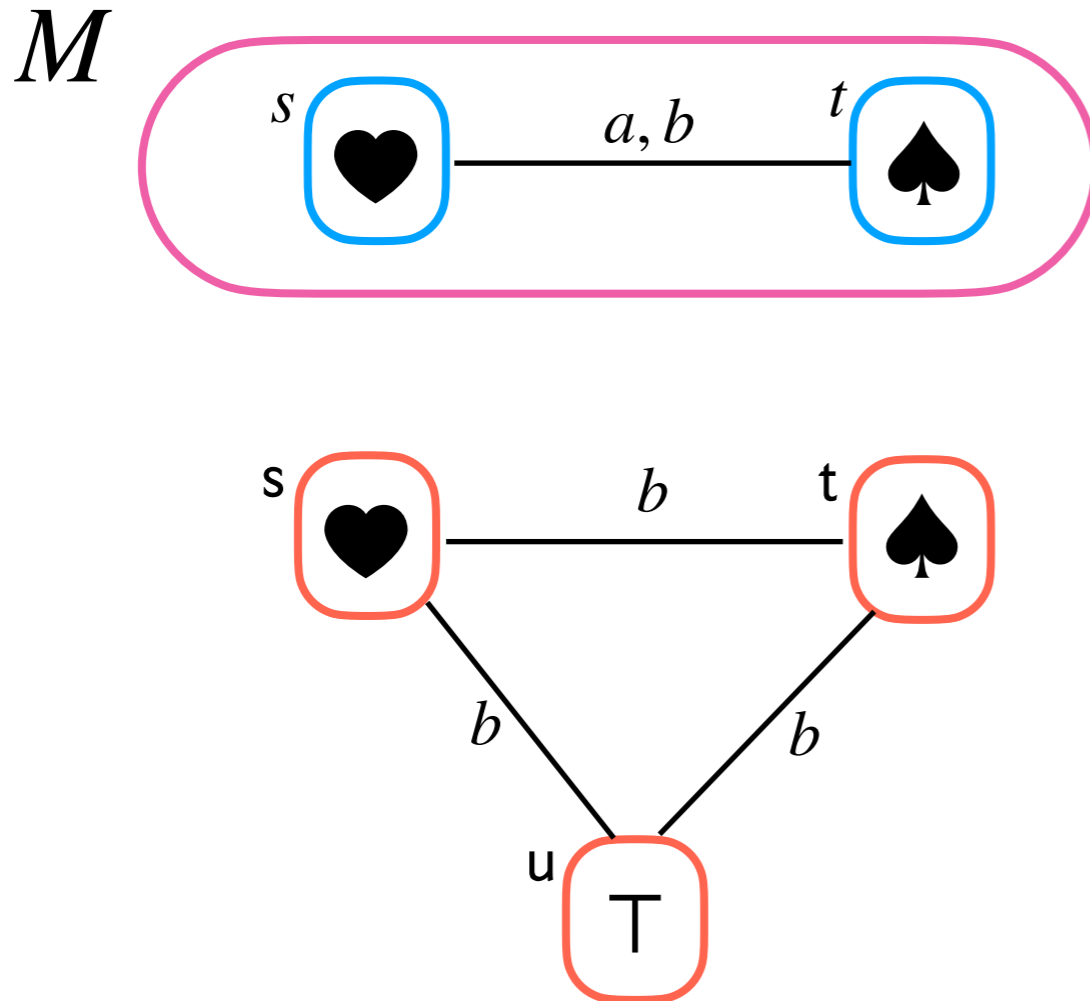
There is a card lying face down on a table that can be either \heartsuit or \spadesuit . Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



Alice could have seen \heartsuit , \spadesuit , or nothing (she did not look)

Card Example

There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



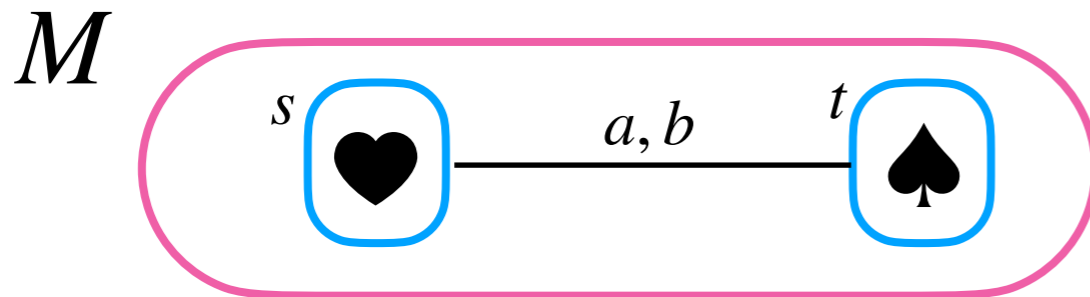
Alice could have seen ♥, ♠, or nothing (she did not look)

And she knows what she did!

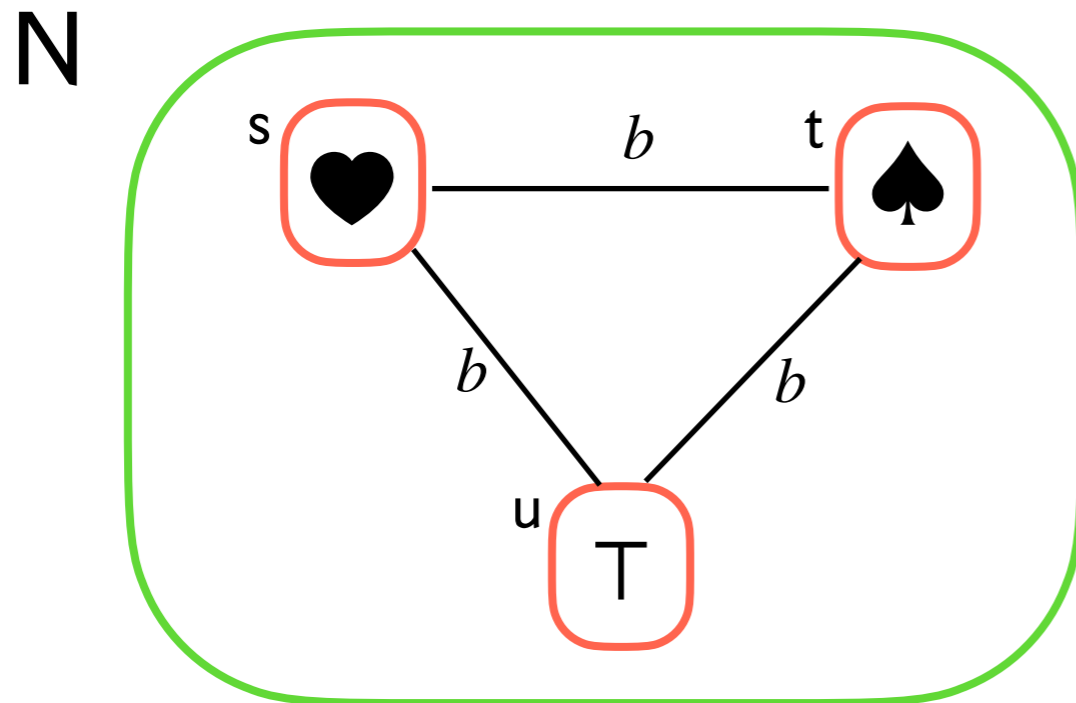
Whereas for Bob, all these opportunities are possible

Card Example

There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



We have something that looks like a model... an **action model!**

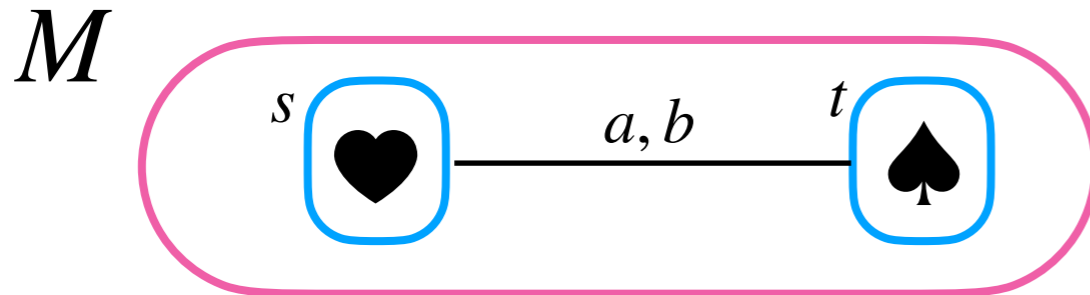


Action models can represent **complex epistemic actions**

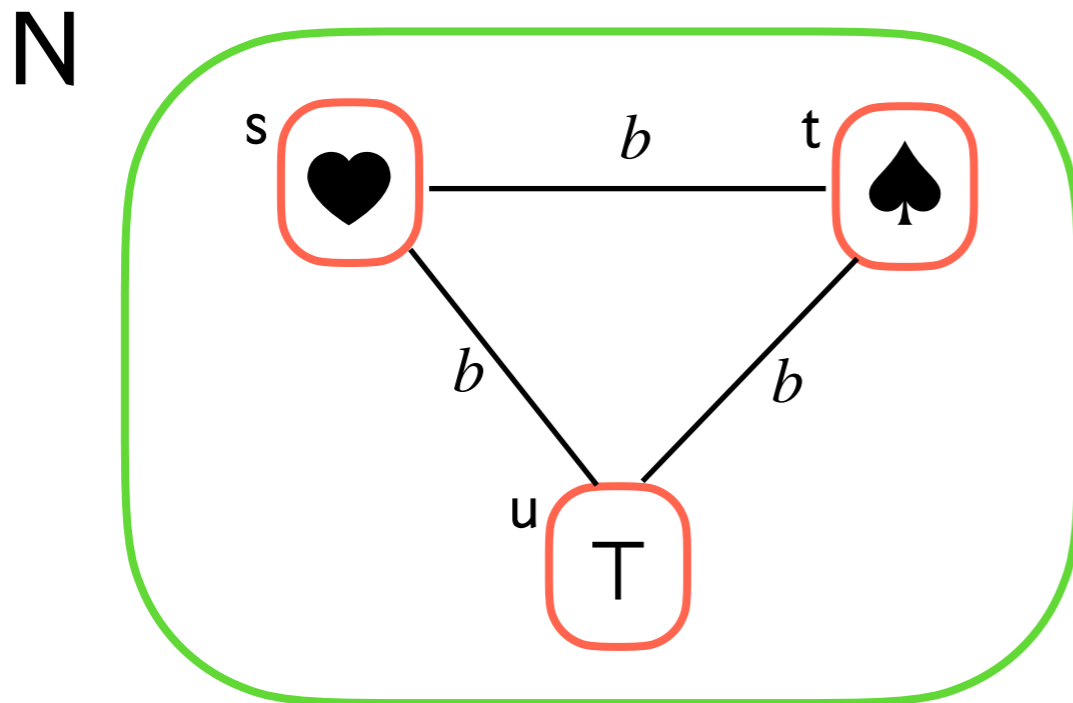
Bob suspects that Alice knows the suit of the card

Card Example

There is a card lying face down on a table that can be either \heartsuit or \spadesuit . Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



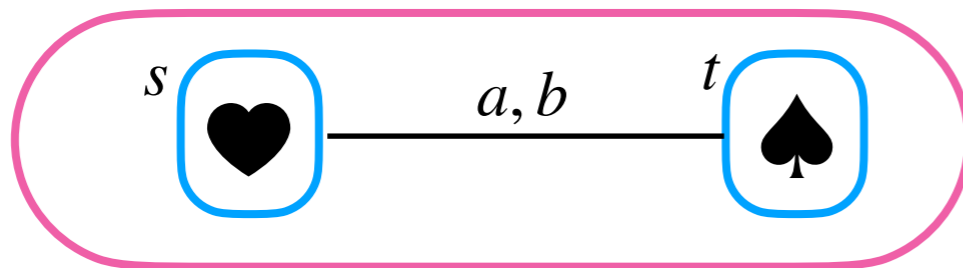
Let's execute action model N in model M



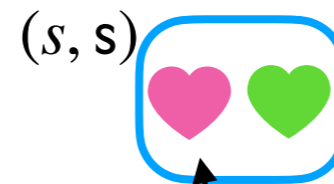
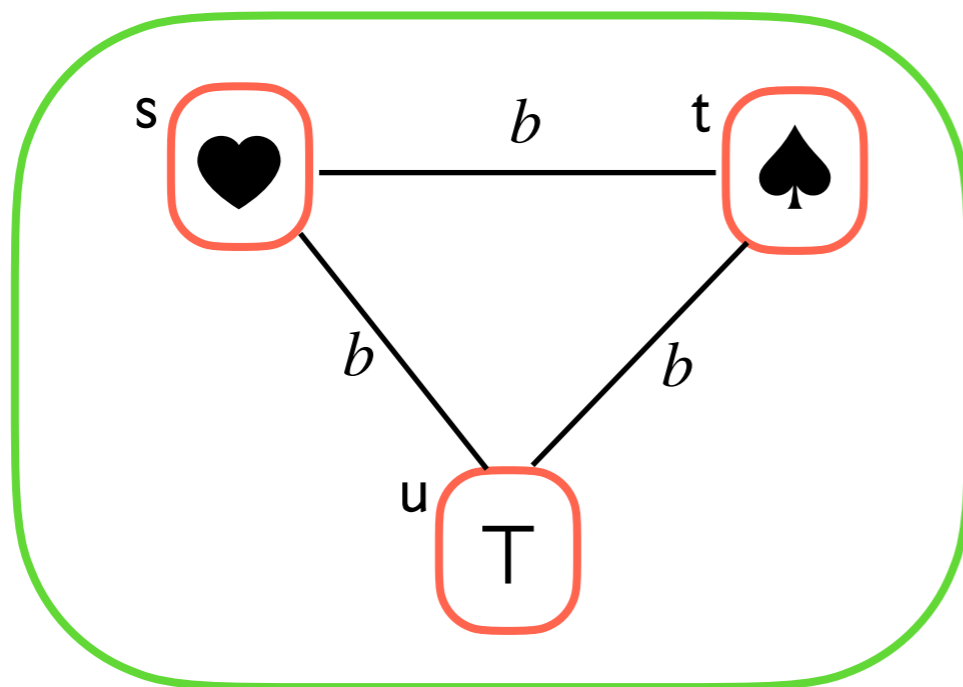
Card Example

There is a card lying face down on a table that can be either \heartsuit or \spadesuit . Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.

M



N

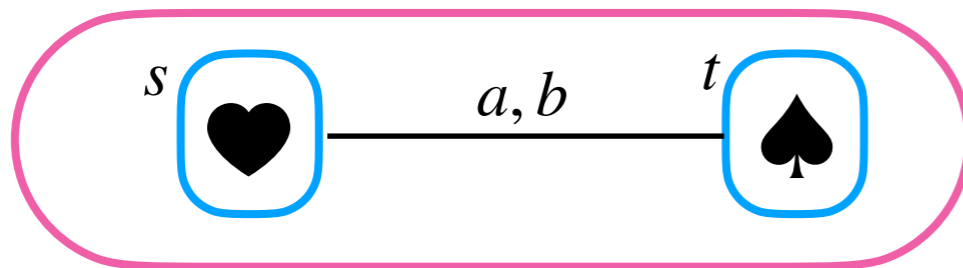


What is the case in the **model**

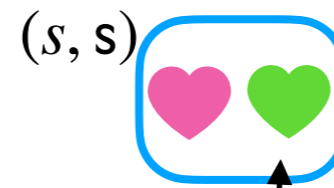
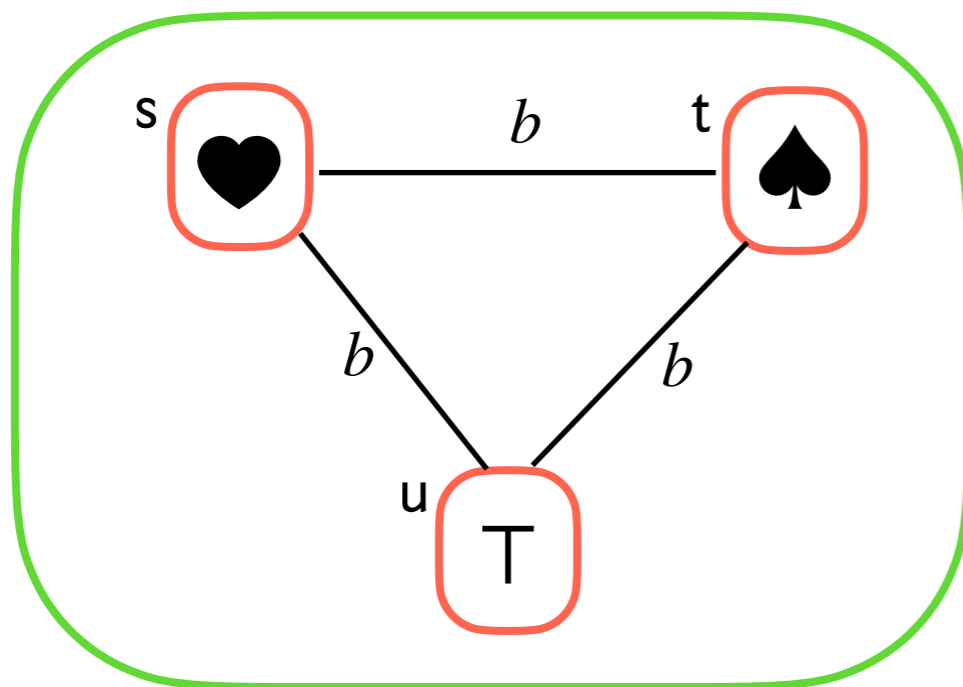
Card Example

There is a card lying face down on a table that can be either \heartsuit or \spadesuit . Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.

M



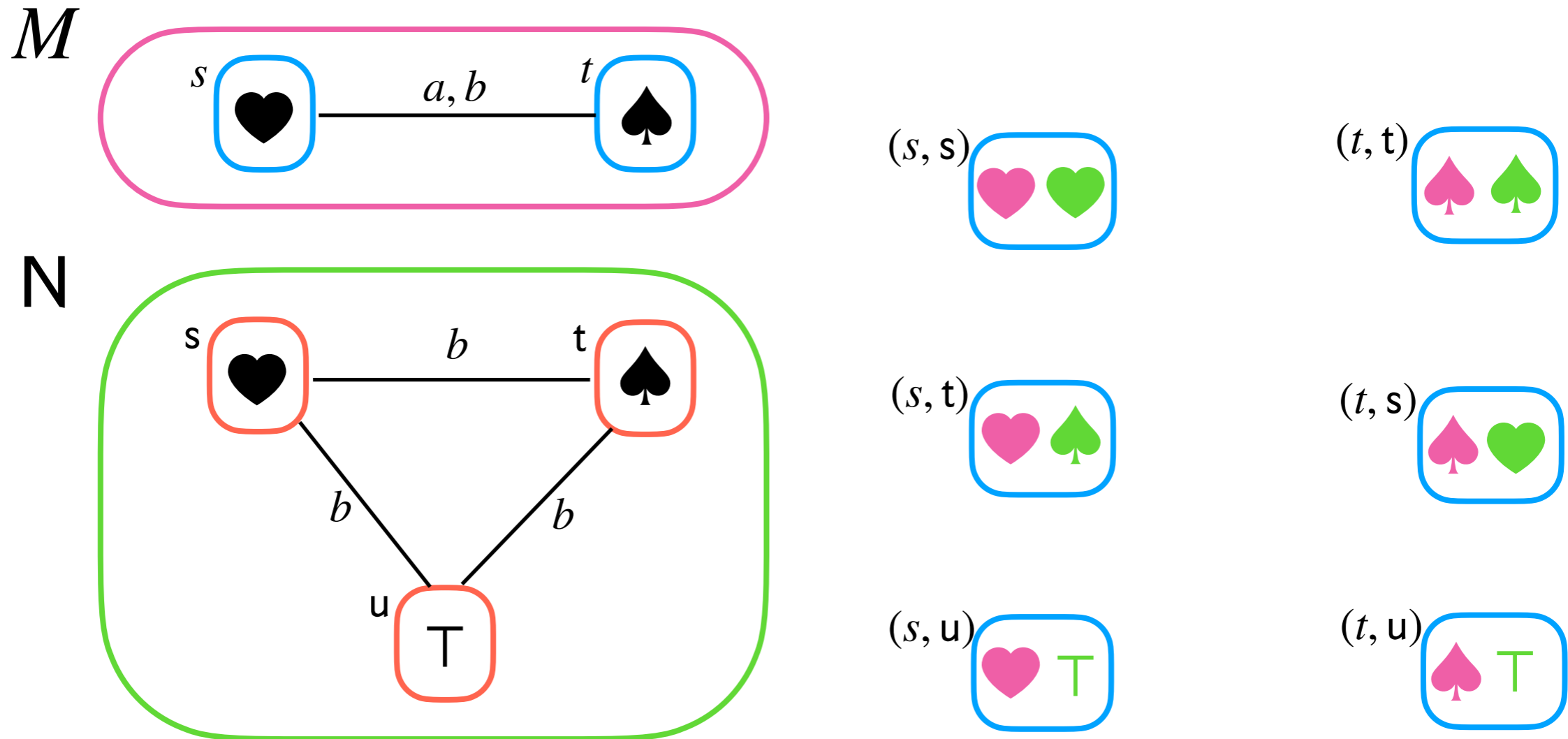
N



What should hold according to the action model

Card Example

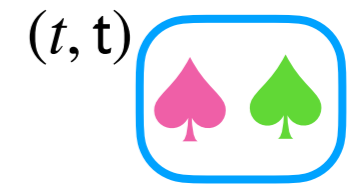
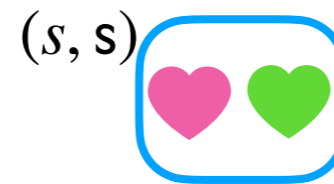
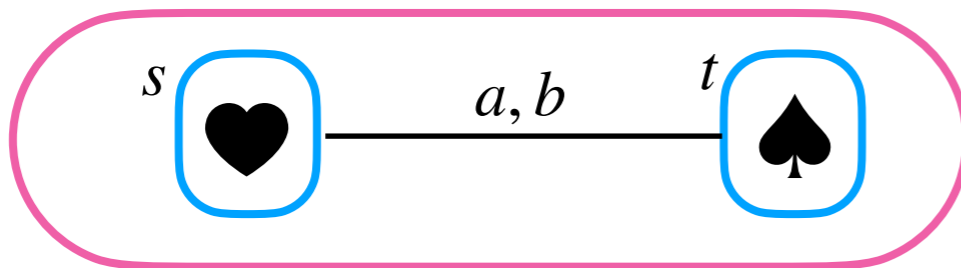
There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



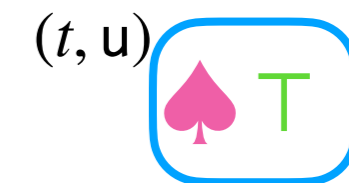
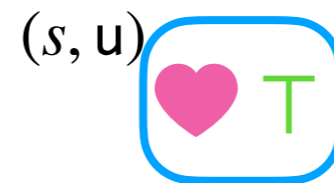
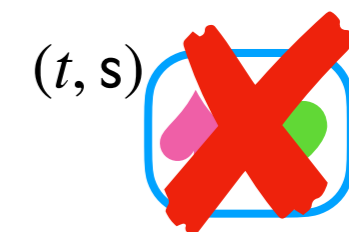
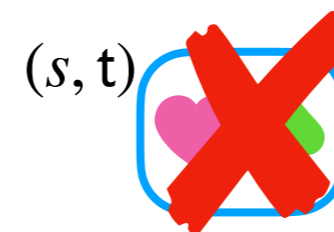
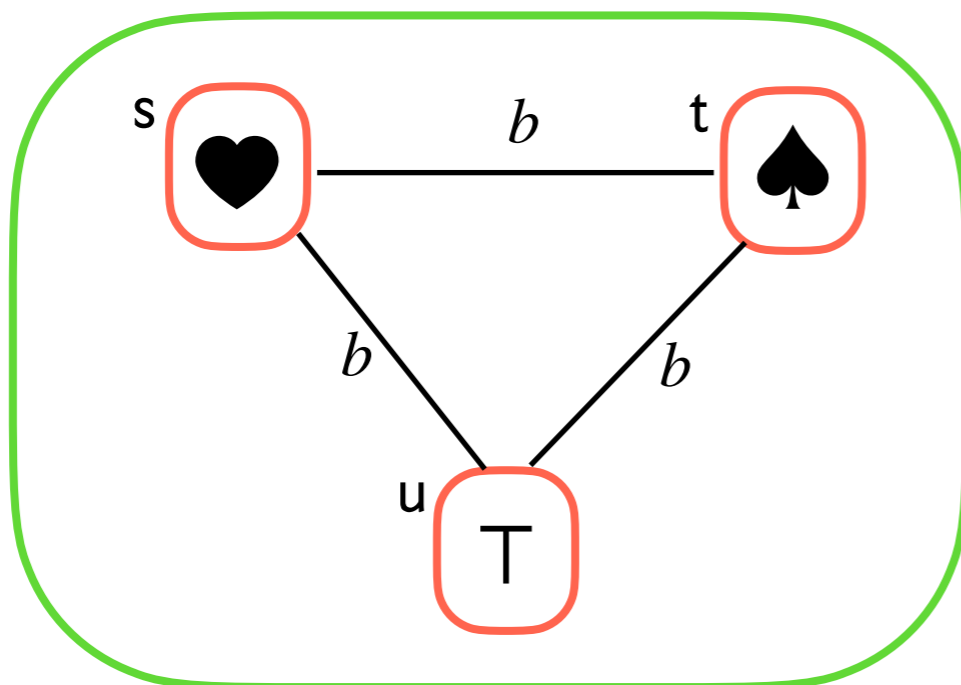
Card Example

There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.

M

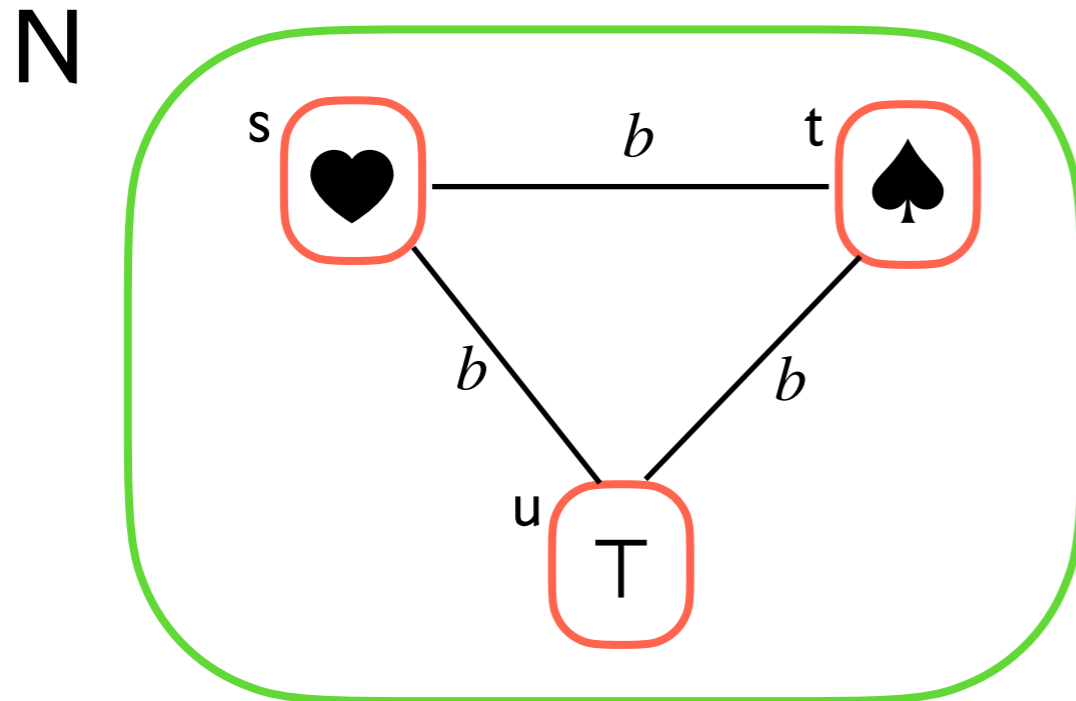
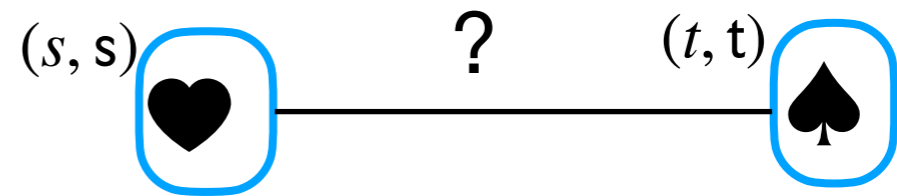
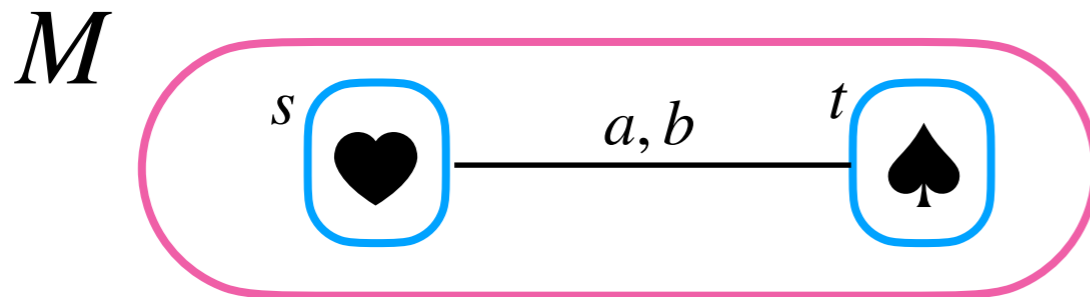


N



Card Example

There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



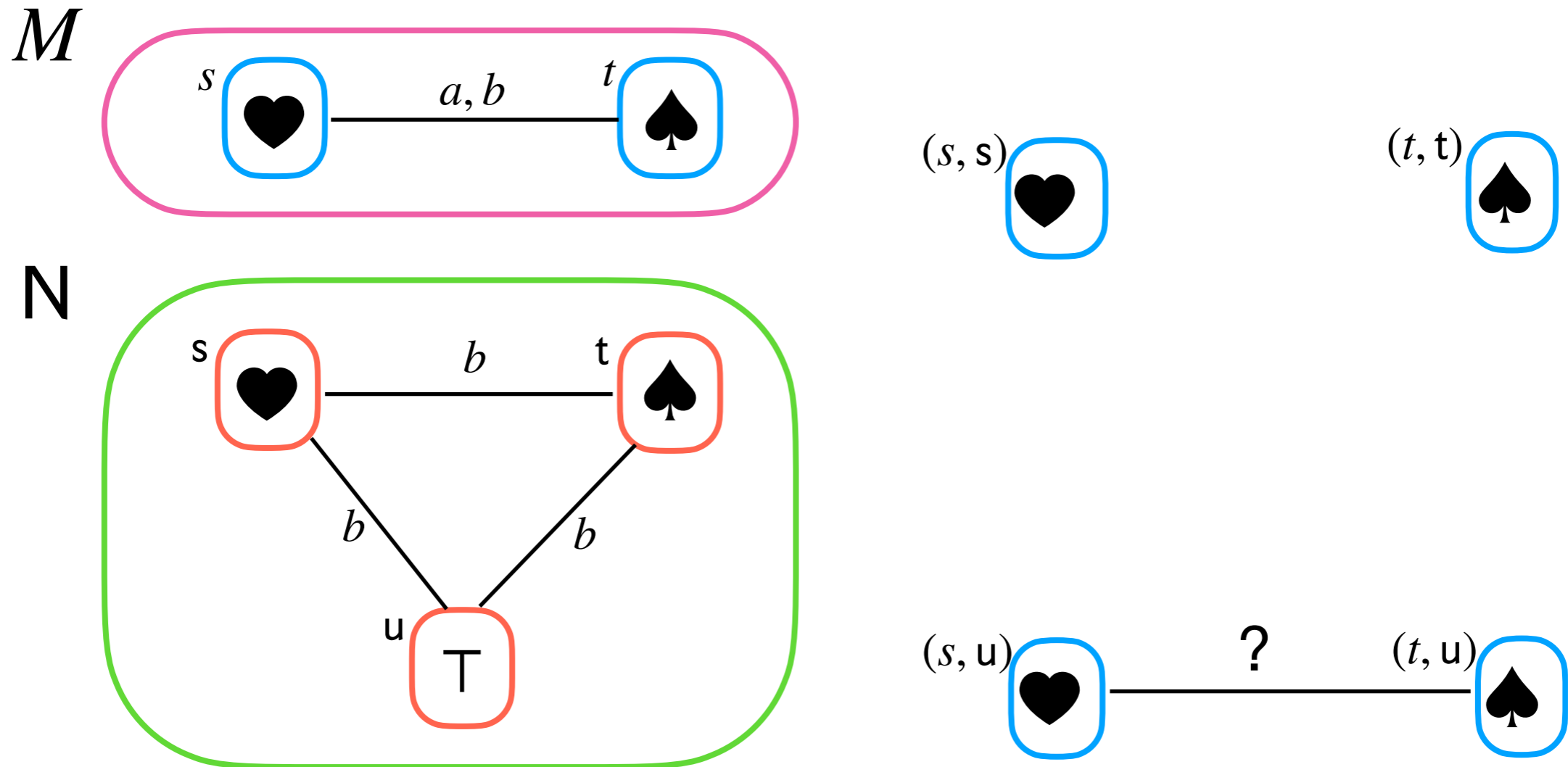
Can Alice distinguish these two outcomes?

What is sufficient for her to distinguish the two states?



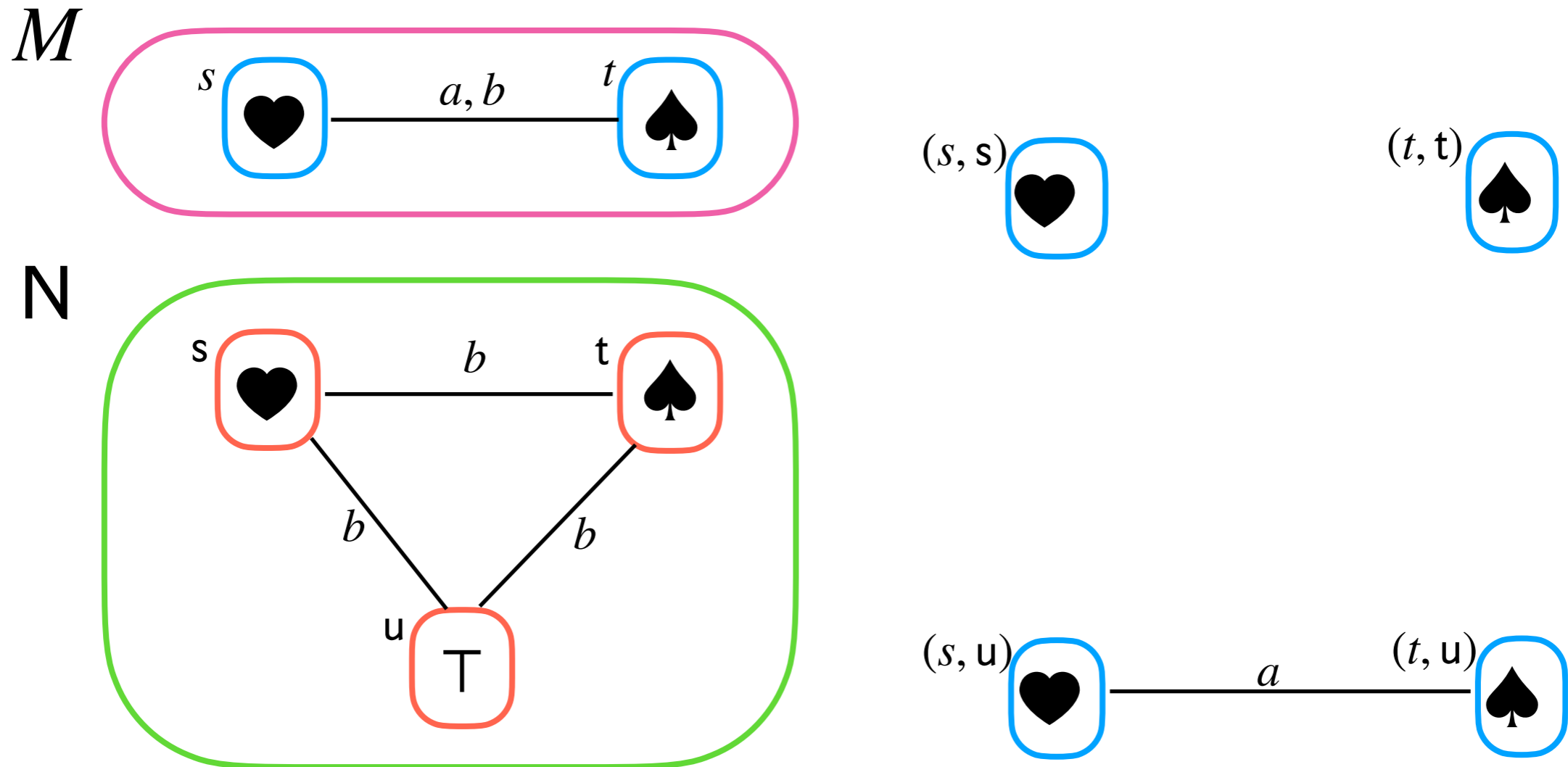
Card Example

There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



Card Example

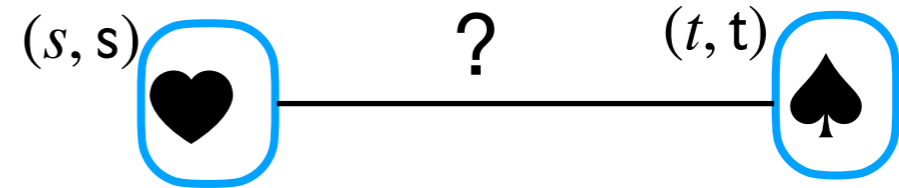
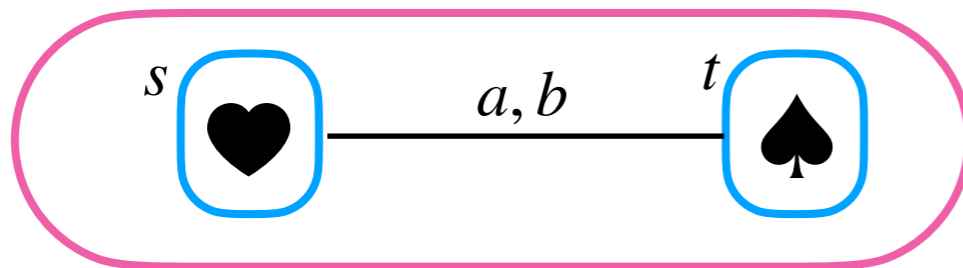
There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



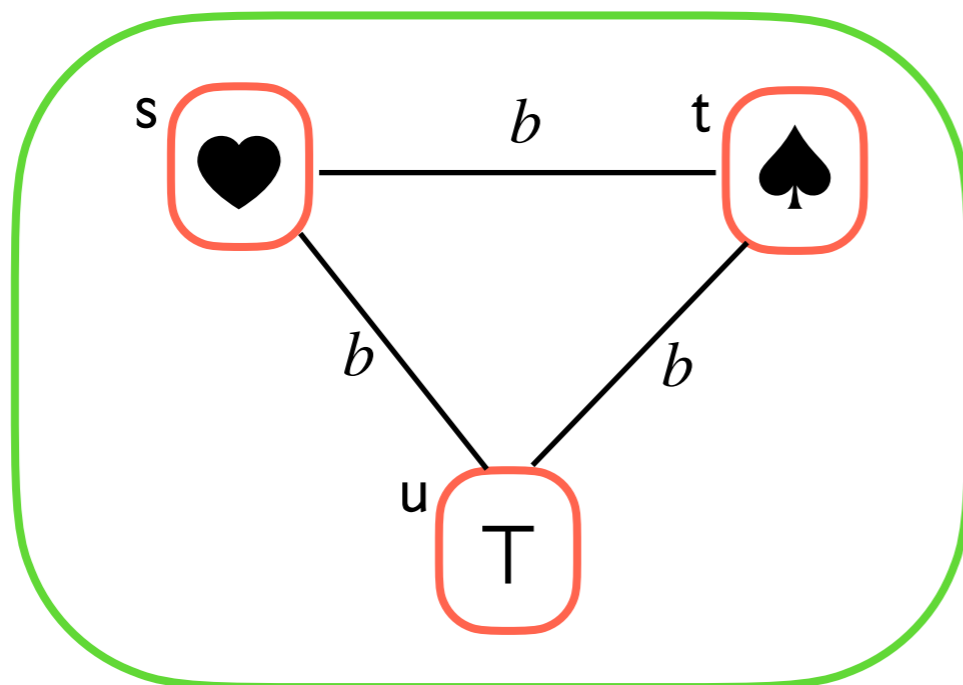
Card Example

There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.

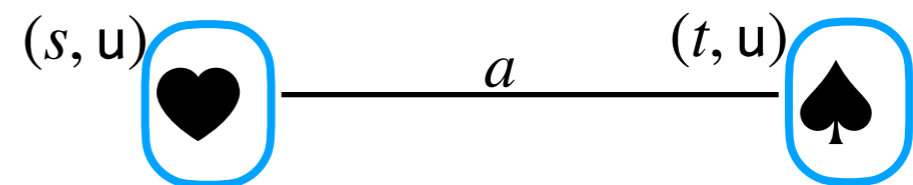
M



N

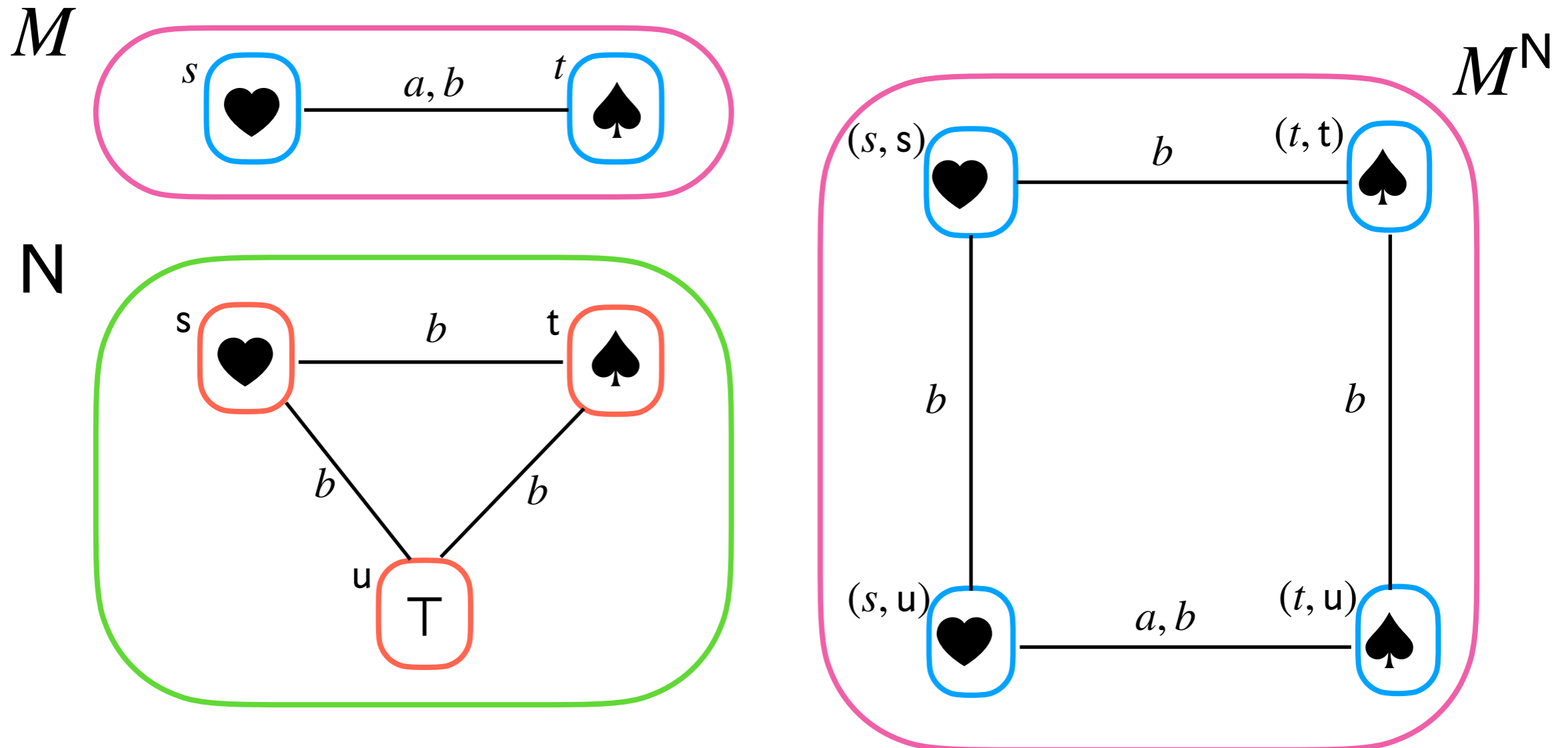


What about Bob?



Card Example

There is a card lying face down on a table that can be either ♥ or ♠. Alice and Bob see the card but do not know its suit. Then Bob walks out, and coming back suspects that Alice have looked at the suit of the card.



Action Model Logic

Language of
AML

$$\mathcal{AML} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi \mid [N_t]\varphi$$

Action model

- An **action model** N is a tuple (S, \sim, pre) , where
- $S \neq \emptyset$ is a set of states;
 - $R : A \rightarrow 2^{S \times S}$ is an indistinguishability function with each R_a being an equivalence relation;
 - $\text{pre} : S \rightarrow \mathcal{L}$ is the precondition function.

Action Model Logic

Language of
AML

$$\mathcal{AML} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi \mid [N_t]\varphi$$

Semantics

$$M_s \models [N_t]\varphi \text{ iff } M_s \models \text{pre}(t) \text{ implies } M_{(s,t)}^N \models \varphi$$

$$M_s \models \langle N_t \rangle \varphi \text{ iff } M_s \models \text{pre}(t) \text{ and } M_{(s,t)}^N \models \varphi$$

Semantics PAL

$$M_s \models [\psi]\varphi \text{ iff } M_s \models \psi \text{ implies } M_s^\psi \models \varphi$$

$$M_s \models \langle \psi \rangle \varphi \text{ iff } M_s \models \psi \text{ and } M_s^\psi \models \varphi$$

Action Model Logic

Language of
AML

$$\mathcal{AML} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi \mid [N_t]\varphi$$

Semantics

$$M_s \models [N_t]\varphi \text{ iff } M_s \models \text{pre}(t) \text{ implies } M_{(s,t)}^N \models \varphi$$

$$M_s \models \langle N_t \rangle \varphi \text{ iff } M_s \models \text{pre}(t) \text{ and } M_{(s,t)}^N \models \varphi$$

Updated model

Let $M = (S, \sim, V)$ and $N = (S, R, \text{pre})$. An **updated model** M^N is a tuple (S^N, \sim^N, V^N) , where

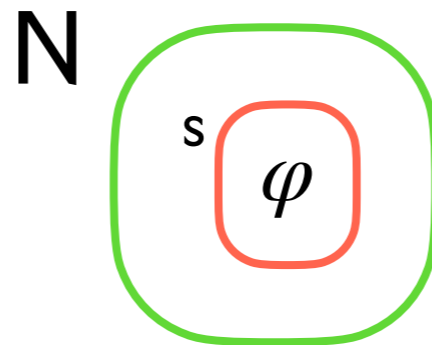
- $S^N = \{(s, t) \mid s \in S, t \in S, M_s \models \text{pre}(t)\}$;
- $(s, t) \sim_a^N (u, v)$ iff $s \sim_a u$ and $tR_a v$;
- $(s, t) \in V^N(p)$ iff $s \in V(p)$.

Overview of AML So Far

- **Action models** allow modelling of plethora of epistemic events
- **Execution** of an action model is done via a cross product with a given epistemic model

What do you think, how do action models stand related to public announcements?

Public
announcement
of φ



$$N = (\{s\}, \\ \{sR_a s \mid a \in A\}, \\ \text{pre}(s) = \varphi)$$

Overview of AML So Far

- **Action models** allow modelling of plethora of epistemic events
- **Execution** of an action model is done via a cross product with a given epistemic model
- Action models **can model public announcements**
- Sooooo....
- How much **expressivity** do we get, compared to the standard EL?

Overview of AML So Far

- **Action models** allow modelling of plethora of epistemic events
- **Execution** of an action model is done via a cross product with a given epistemic model
- Action models **can model public announcements**
- Sooooo....
- How much **expressivity** do we get, compared to the standard EL? Again, **none at all!**

Axiomatisation of AML

Axioms of EL

$$[N_t]p \leftrightarrow (\text{pre}(t) \rightarrow p)$$

$$[N_t]\neg\psi \leftrightarrow (\text{pre}(t) \rightarrow \neg[N_t]\psi)$$

$$[N_t](\psi \wedge \chi) \leftrightarrow ([N_t]\psi \wedge [N_t]\chi)$$

$$[N_t]\Box_a\psi \leftrightarrow$$

$$\leftrightarrow (\text{pre}(t) \rightarrow \bigwedge_{tR_a u} \Box_a [N_u]\psi)$$

$$[N_t][O_u]\psi \leftrightarrow [N_t; O_u]\psi$$

From φ infer $[N_t]\psi$

Theorem. AML and EL are equally expressive

Theorem. AML is sound and complete

Theorem. Complexity of SAT-AML is NEXPTIME-complete

Theorem. Complexity of MC-AML is PSPACE-complete

Actions Models vs. Public Announcements

So, both AML and PAL are as expressive as EL via reduction axioms

But **action models seem more expressive** than public announcements...

And they indeed are! In a way...

On the one hand, we saw that for each public announcement there is an action model that results in the same updated model

On the other hand, action models can make the updated model bigger than the original one (which announcements cannot do)

Thus...

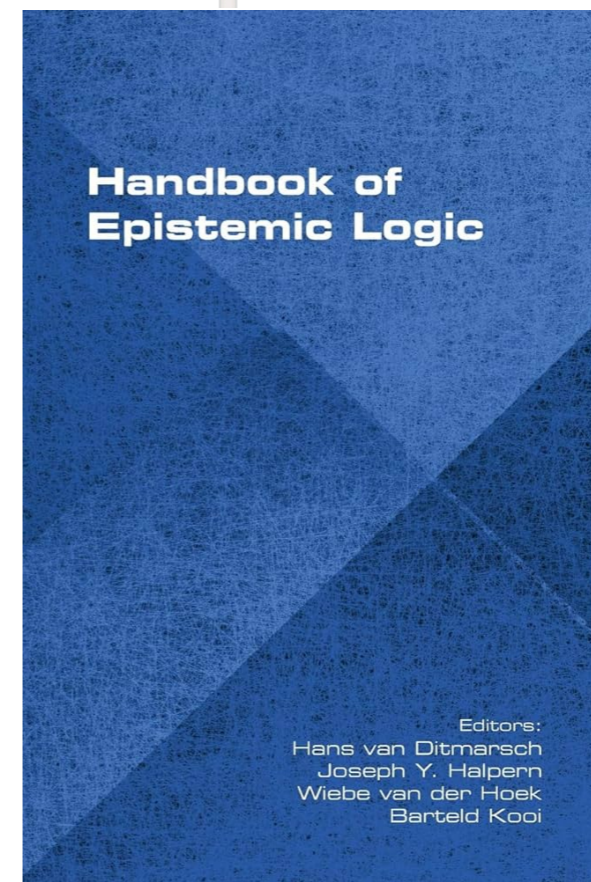
Actions Models vs. Public Announcements

Theorem. Update expressivity of AML is strictly greater than that of PAL

Beyond Announcements and Action Models

- PAL and AML are but only two representatives of DELs. We can have so much more!
- Ontic changes
- Adding and removing arrows
- Communication within groups of agents
- Everything above in the context of group knowledge
- And so on and so on and so on and so on...

Where To Start



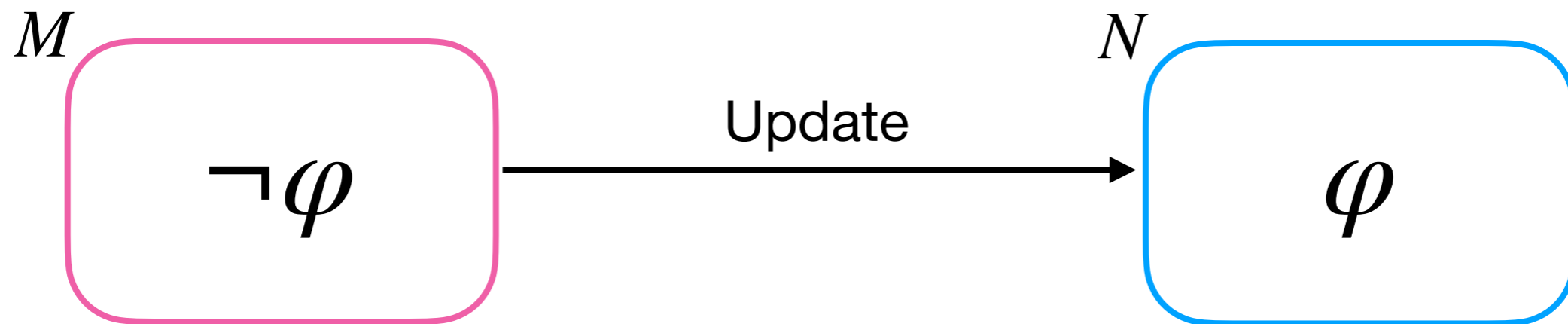
Part IV

Current Research Directions

I. Quantification in DEL

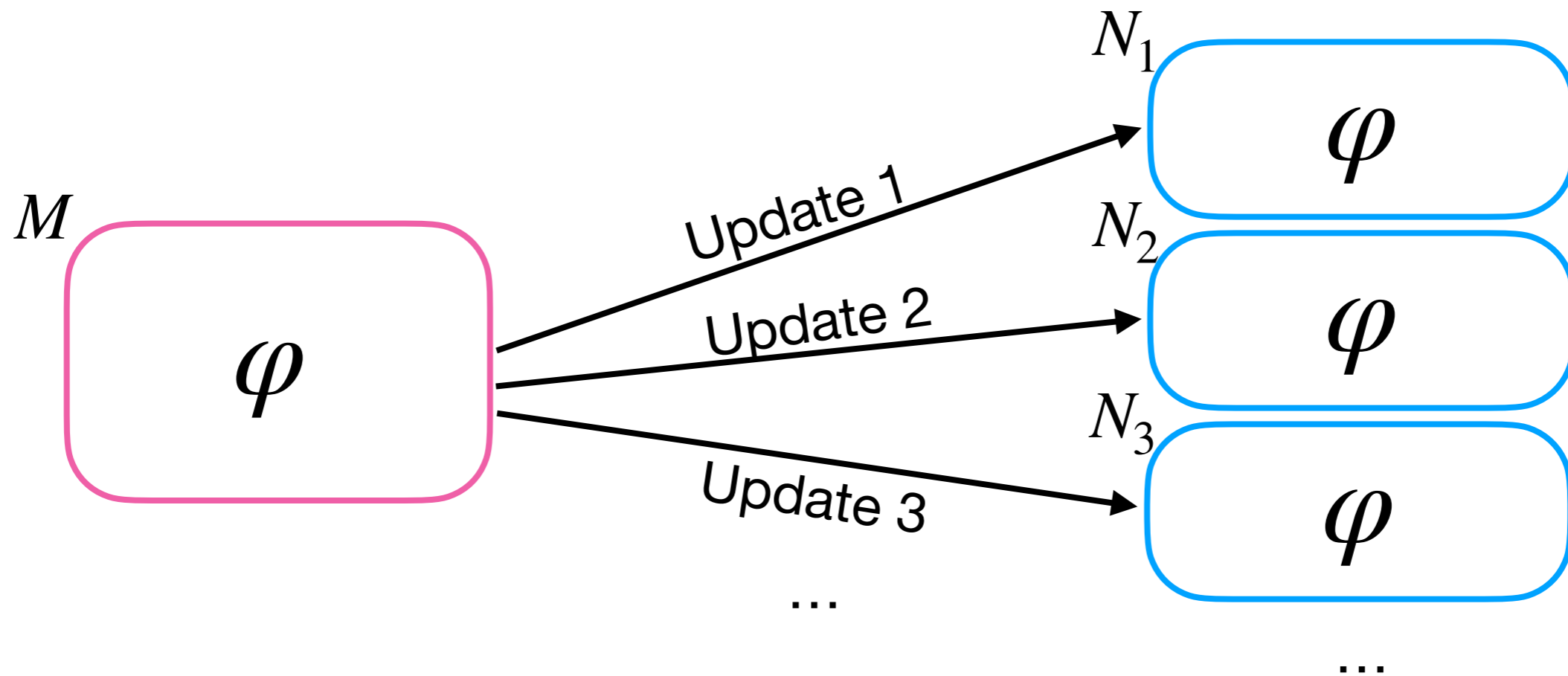
II. Theory of Mind

Quantifying Over Updates



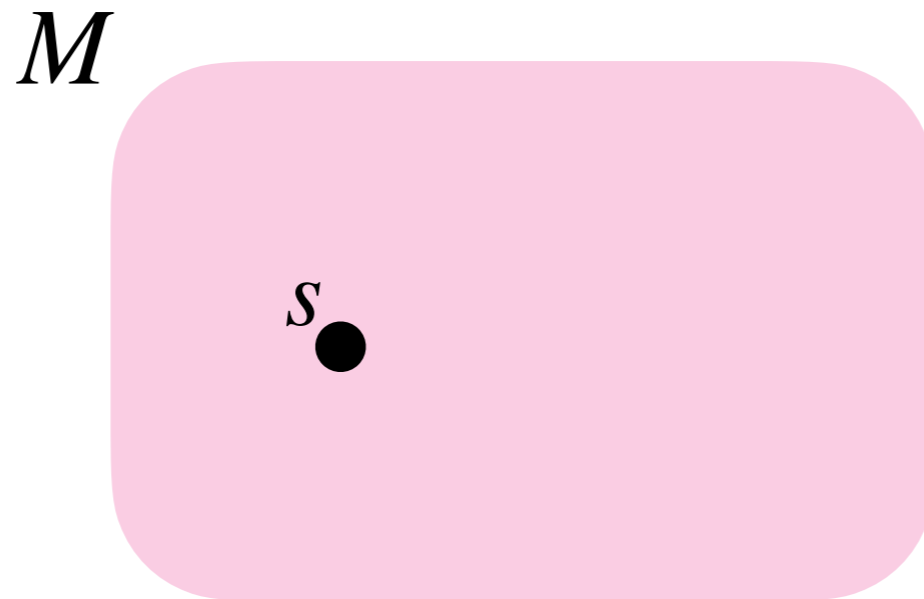
Existence: Having a starting configuration M and a property φ we would like to have, **there is an epistemic action** that results in configuration N satisfying φ

Quantifying Over Updates



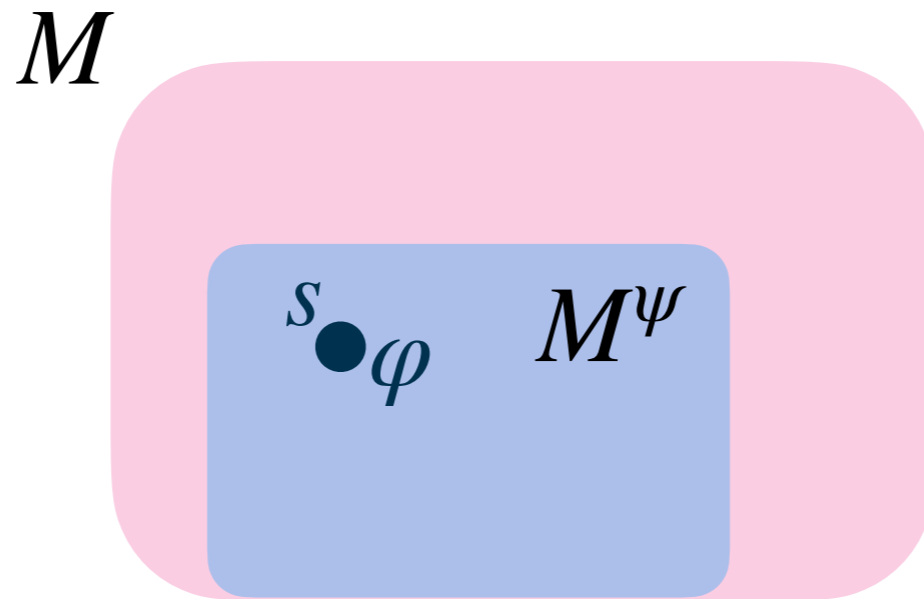
Universality: Having a starting configuration M satisfying φ , we would like to ensure that **all epistemic actions** result in some configuration N satisfying φ

Quantifying Over Public Announcements



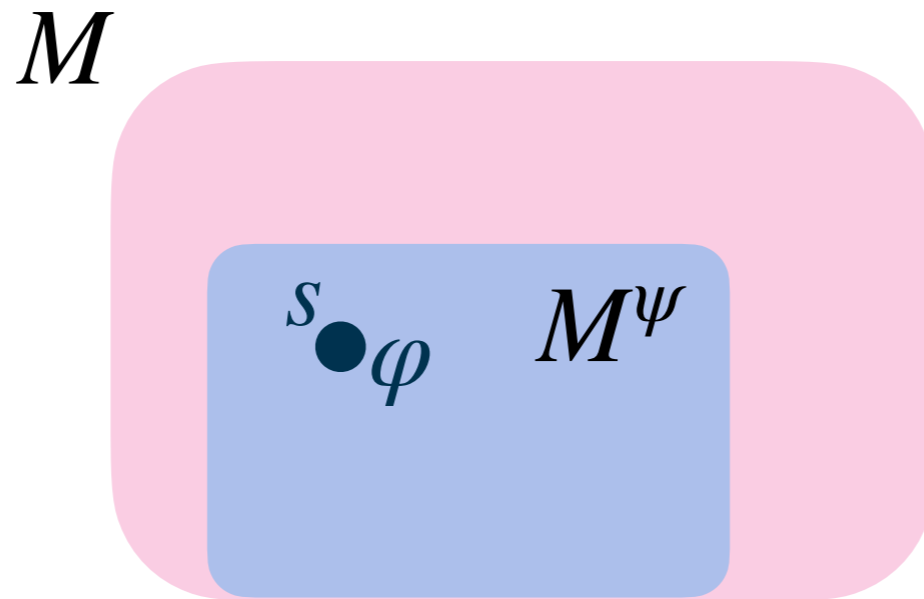
$\langle ! \rangle \varphi$: There is a public announcement, after which φ is true

Quantifying Over Public Announcements



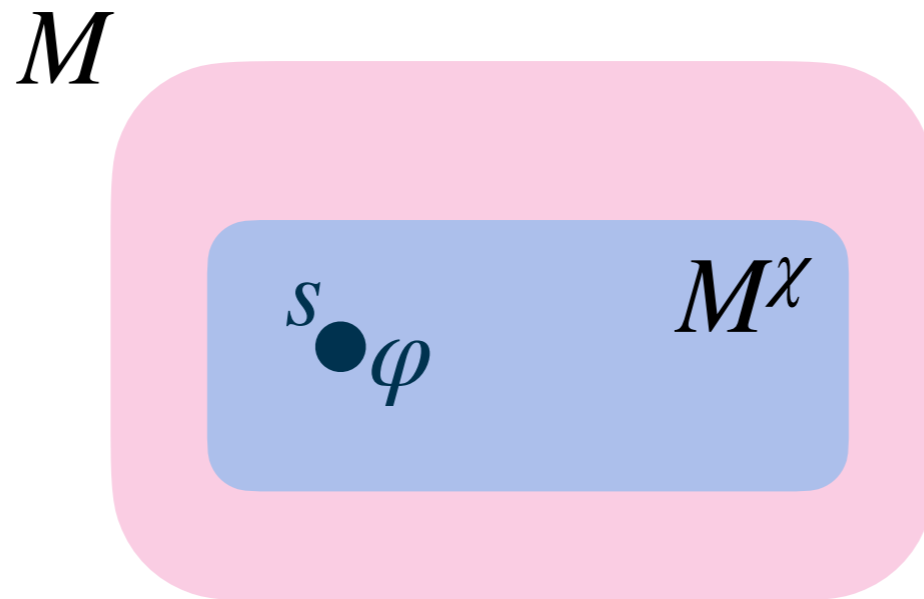
$\langle ! \rangle \varphi$: There is a public announcement, after which φ is true

Quantifying Over Public Announcements



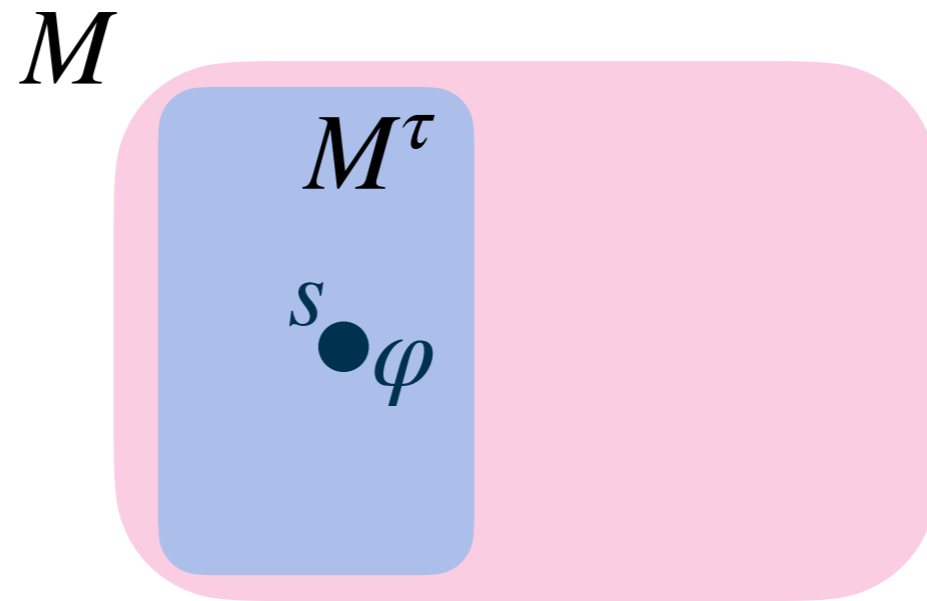
$[!]\varphi$: After all public announcements, φ is true

Quantifying Over Public Announcements



$[!]\varphi$: After all public announcements, φ is true

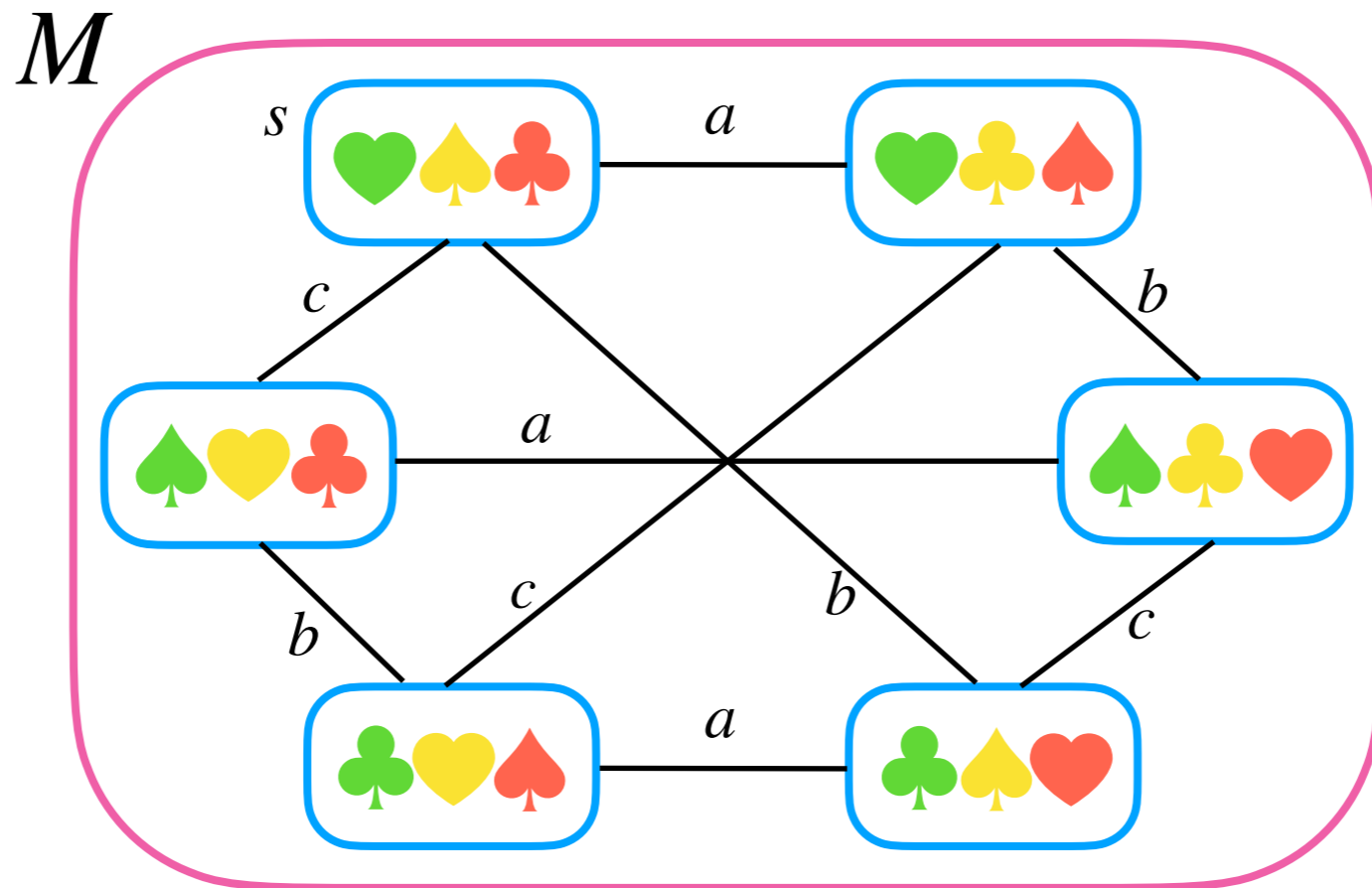
Quantifying Over Public Announcements



$[!]\varphi$: After all public announcements, φ is true

Card Example

There is an announcement such that **Alice** knows the deal, and **Bob** and **Carol** do not

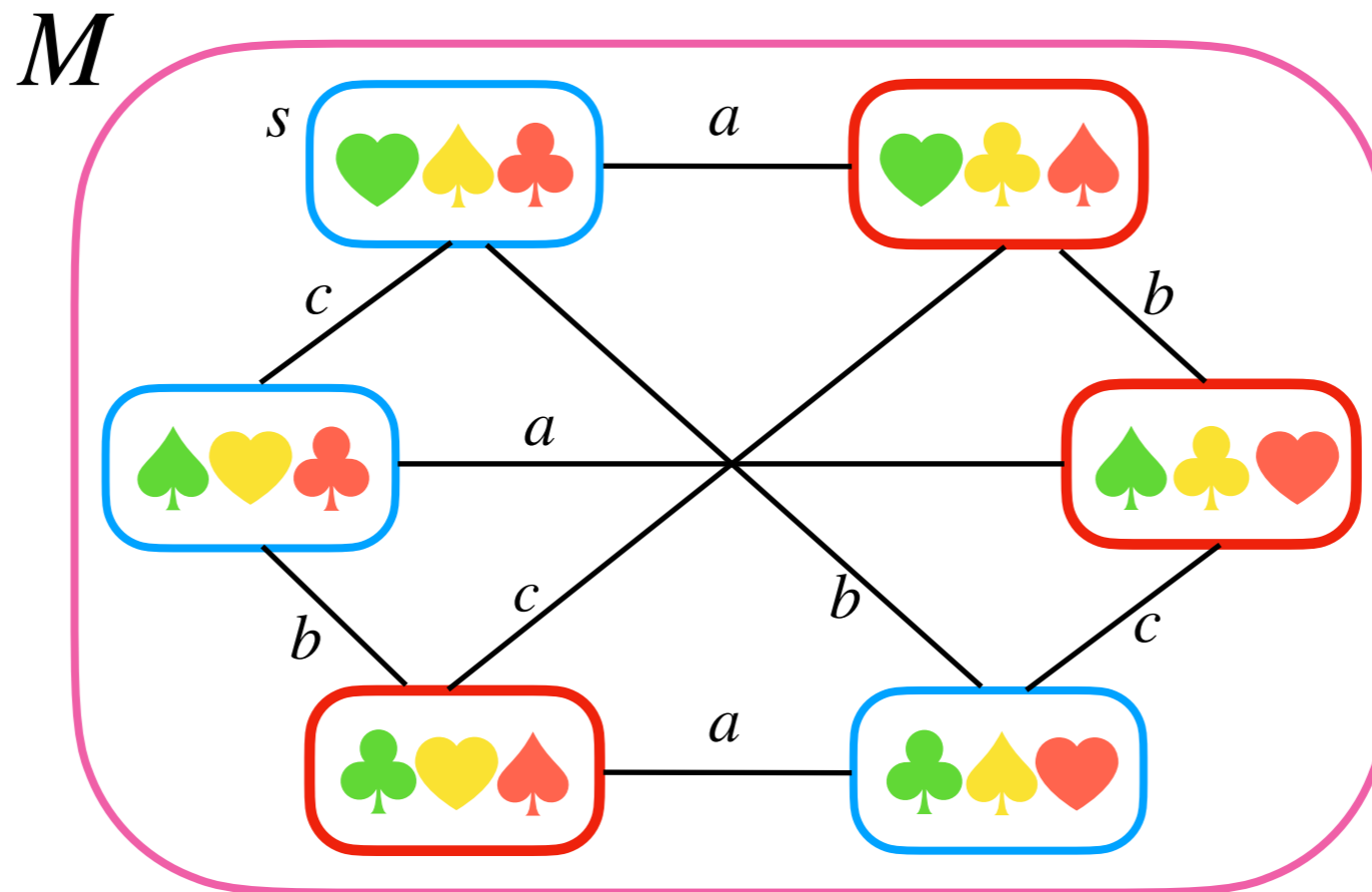


$$M_s \models \langle ! \rangle (\Box_a \text{deal} \wedge \neg \Box_b \text{deal} \wedge \neg \Box_c \text{deal})$$

$$\varphi := (\spadesuit_b \vee \heartsuit_b) \wedge (\clubsuit_c \vee \heartsuit_c)$$

Card Example

There is an announcement such that **Alice** knows the deal, and **Bob** and **Carol** do not

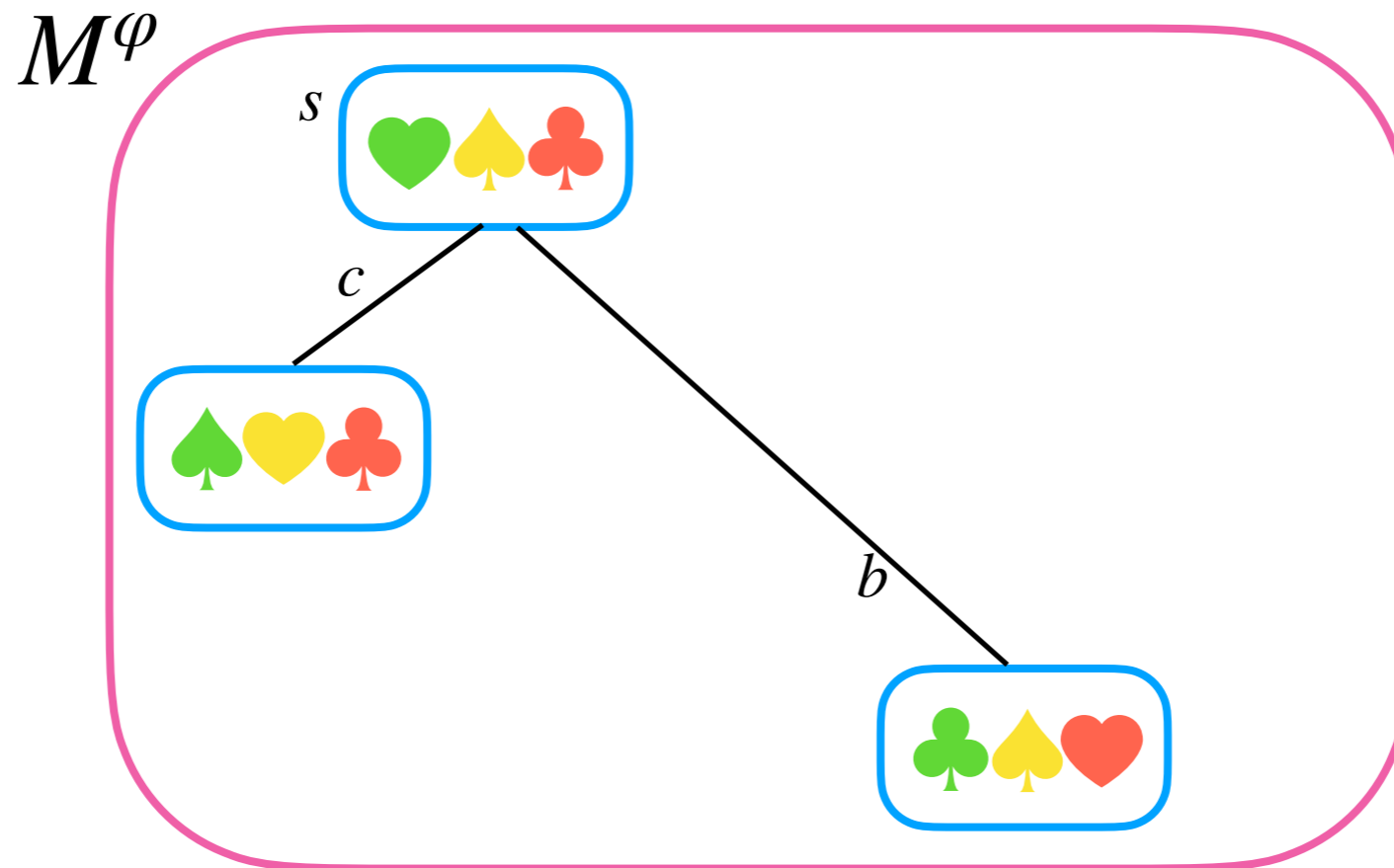


$$M_s \models \langle ! \rangle (\Box_a \text{deal} \wedge \neg \Box_b \text{deal} \wedge \neg \Box_c \text{deal})$$

$$\varphi := (\spadesuit_b \vee \heartsuit_b) \wedge (\clubsuit_c \vee \heartsuit_c)$$

Card Example

There is an announcement such that **Alice** knows the deal,
and **Bob** and **Carol** do not

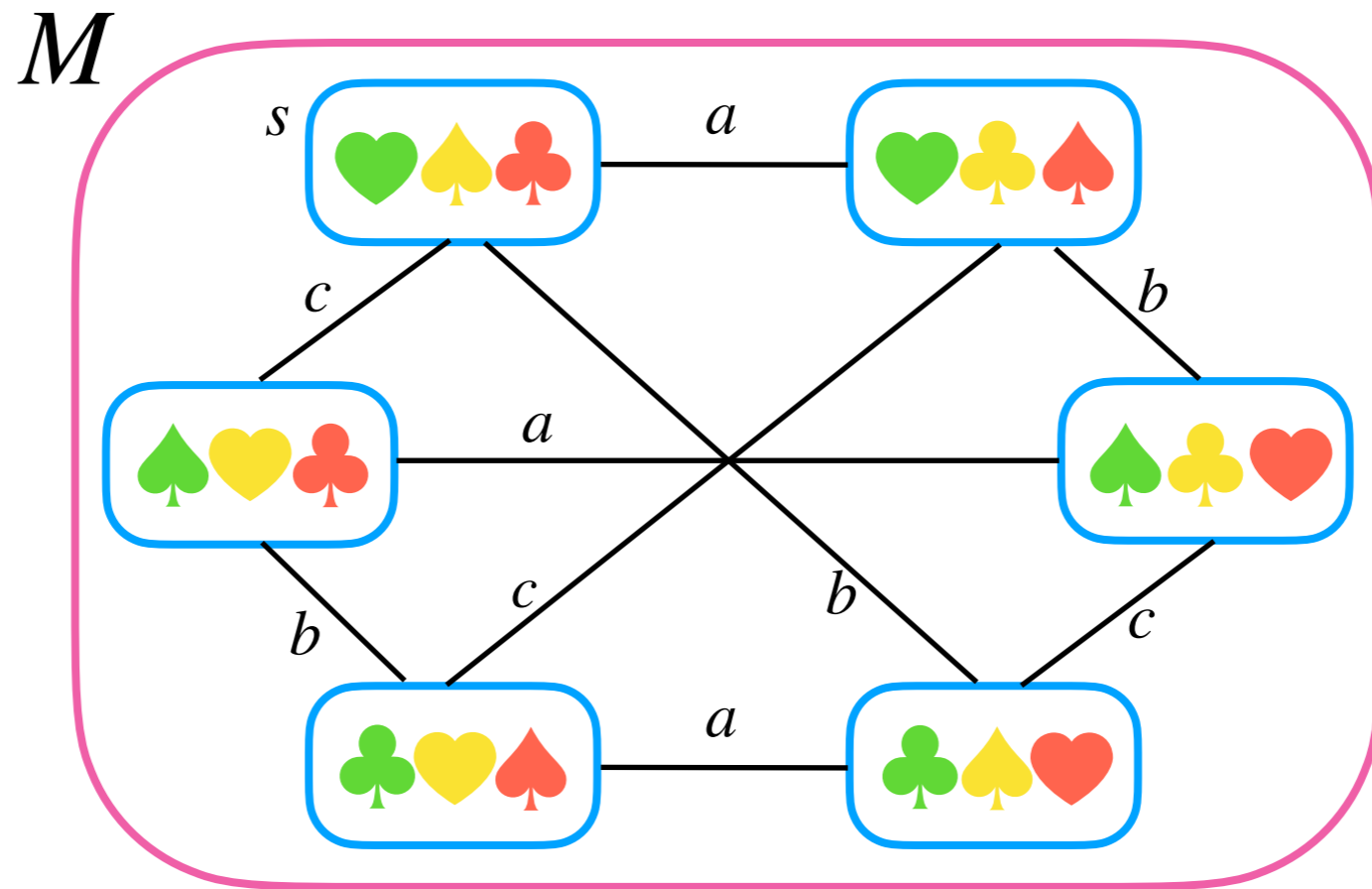


$$M_s \models \langle ! \rangle (\Box_a \text{deal} \wedge \neg \Box_b \text{deal} \wedge \neg \Box_c \text{deal})$$

$$\varphi := (\spadesuit_b \vee \heartsuit_b) \wedge (\clubsuit_c \vee \heartsuit_c)$$

Card Example

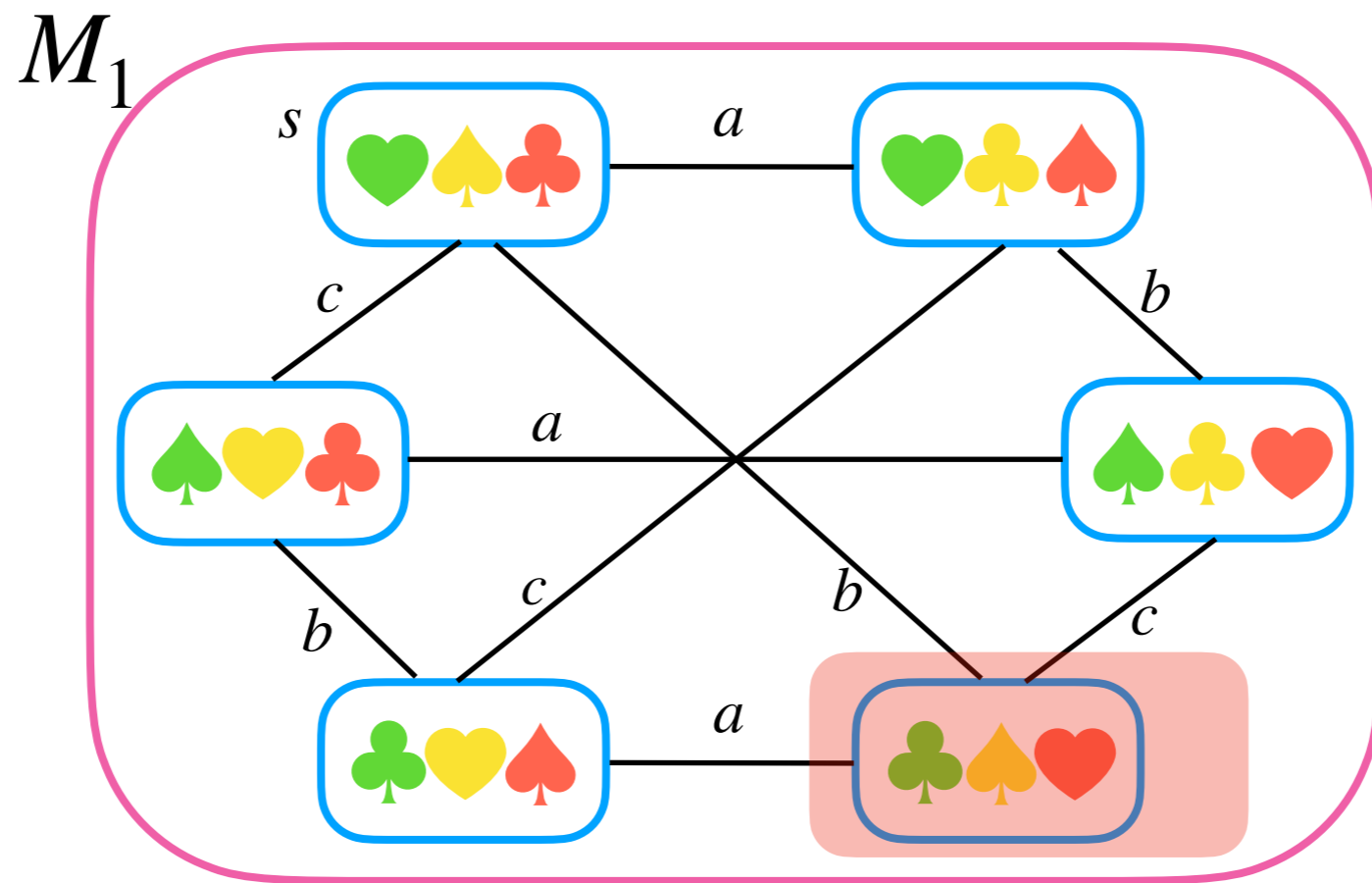
After any announcement, Alice has one of the cards



$$M_s \models [!](\heartsuit_a \vee \clubsuit_a \vee \spadesuit_a)$$

Card Example

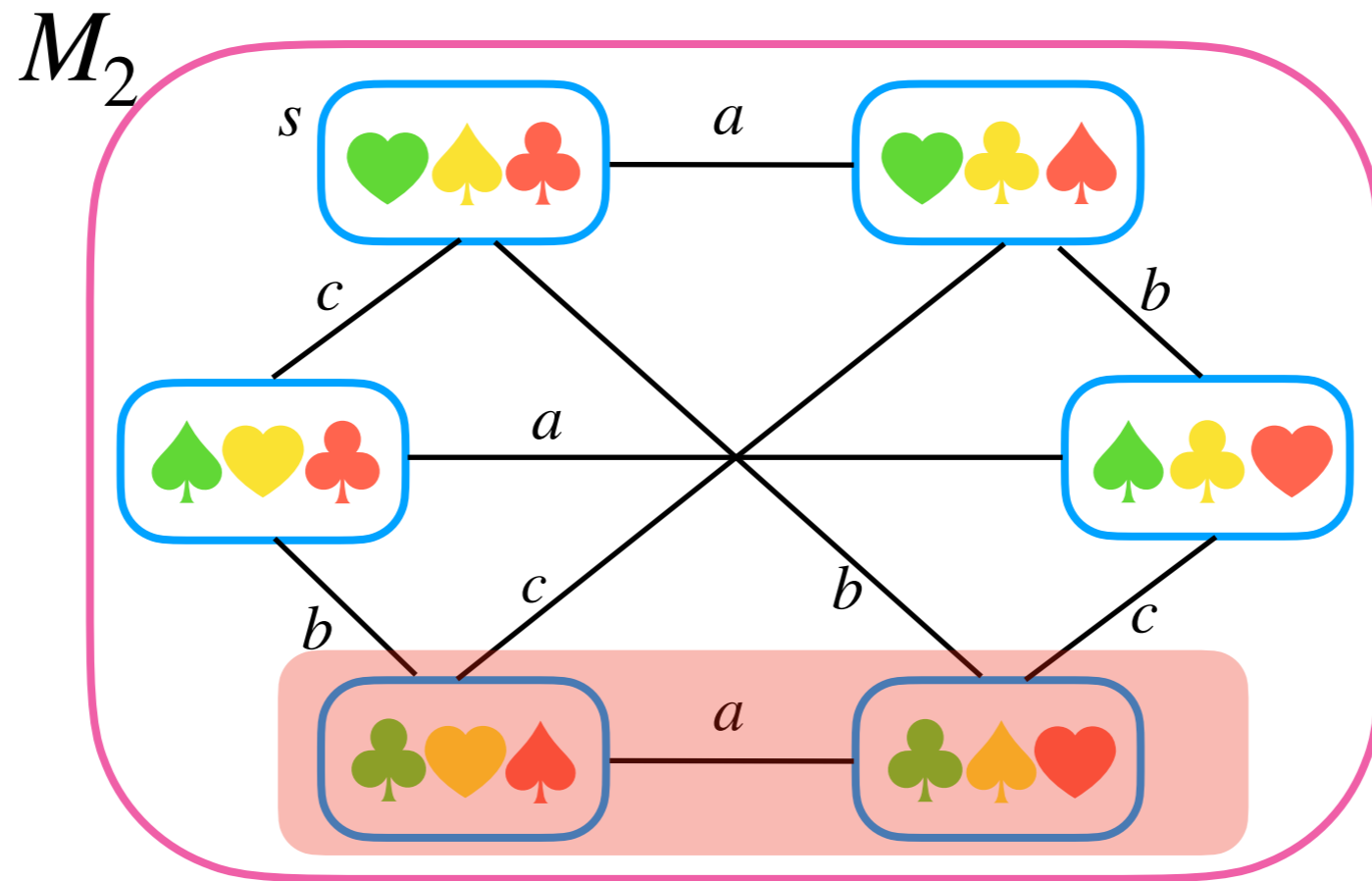
After any announcement, **Alice** has one of the cards



$$M_s \models [!](\heartsuit_a \vee \clubsuit_a \vee \spadesuit_a)$$

Card Example

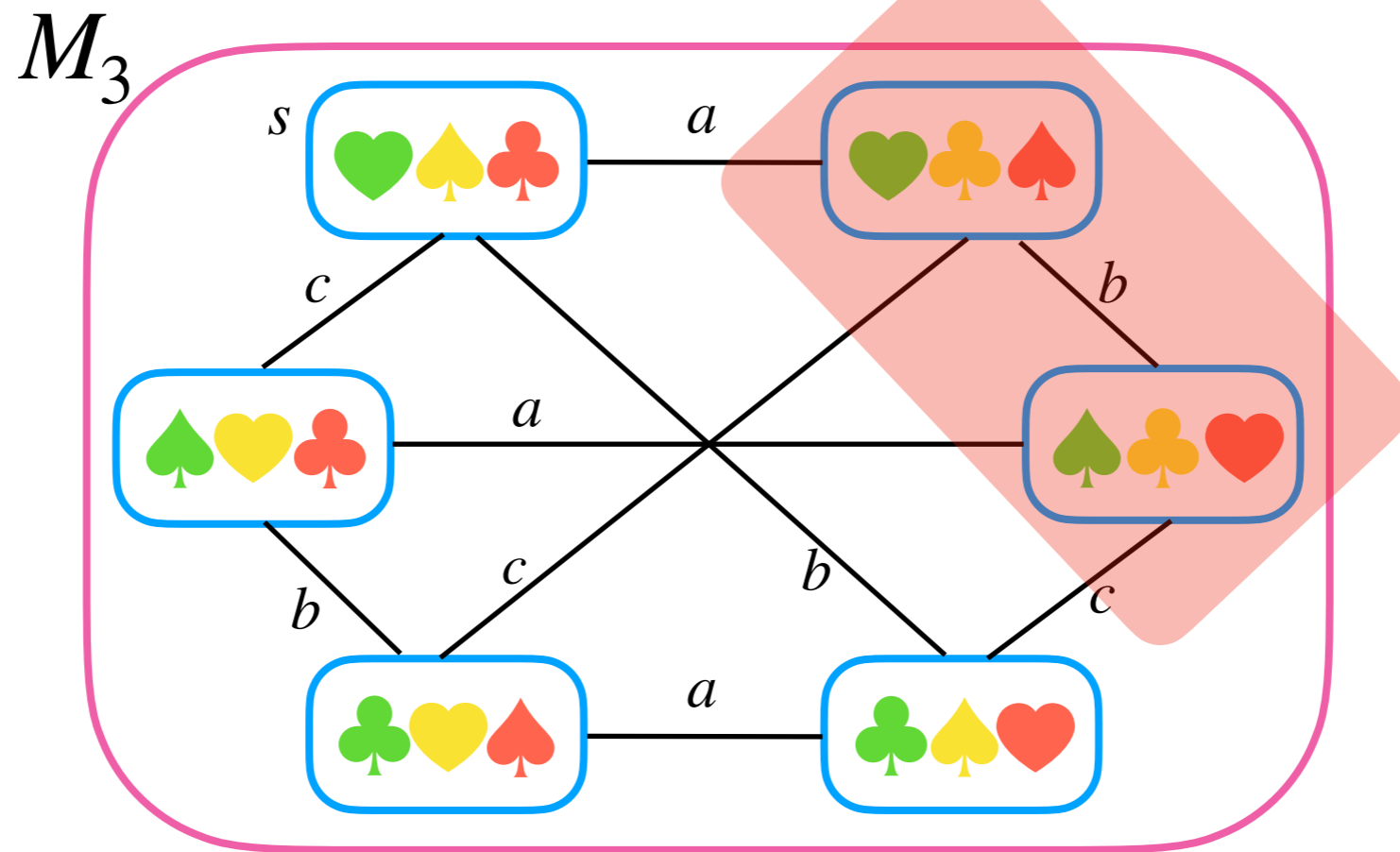
After any announcement, Alice has one of the cards



$$M_s \models [!](\heartsuit_a \vee \clubsuit_a \vee \spadesuit_a)$$

Card Example

After any announcement, Alice has one of the cards



$$M_s \models [!](\heartsuit_a \vee \clubsuit_a \vee \spadesuit_a)$$

Arbitrary PAL

Language of
APAL

$$\mathcal{APAL} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi \mid [\varphi]\varphi \mid [!]\varphi$$

Semantics

$$M_s \models [!]\varphi \text{ iff } \forall \psi \in \mathcal{PAL} : M_s \models [\psi]\varphi$$

$$M_s \models \langle ! \rangle \varphi \text{ iff } \exists \psi \in \mathcal{PAL} : M_s \models \langle \psi \rangle \varphi$$

Some validities

$$\langle \psi \rangle \varphi \rightarrow \langle ! \rangle \varphi \quad [!]\varphi \rightarrow \varphi$$

$$\langle ! \rangle \varphi \leftrightarrow \langle ! \rangle \langle ! \rangle \varphi \quad \langle ! \rangle [!]\varphi \leftrightarrow [!]\langle ! \rangle \varphi$$

Quantification is restricted to formulas of PAL in order to avoid circularity

Why Quantification in DEL?

- Verification of **functionality** and **security** of a system

Functionality. There is a protocol that allows agents to achieve their goals

Why Quantification in DEL?

- Verification of **functionality** and **security** of a system

Security. No matter what agents do, they cannot reach some undesirable state

Why Quantification in DEL?

- Verification of **functionality** and **security** of a system
- Epistemic **planning**

Epistemic planning. Given a set of allowed actions, agents are able to construct and execute a plan based on these actions

Why Quantification in DEL?

- Verification of **functionality** and **security** of a system
- Epistemic **planning**
- Protocol **synthesis**

Protocol synthesis. Given a goal state, provide an action (or their sequence), that takes any give state to the goal one

Why Quantification in DEL?

- Verification of **functionality** and **security** of a system
- Epistemic **planning**
- Protocol **synthesis**
- Capturing the notion of **knowability** in philosophy

Knowability. Every true statement is knowable, in principle

Why Quantification in DEL?

- Verification of **functionality** and **security** of a system
- Epistemic **planning**
- Protocol **synthesis**
- Capturing the notion of **knowability** in philosophy
- And so on and so on and so on and so on...

Knowability. Every true statement is knowable, in principle

APAL versus PAL

Theorem. PAL and EL are equally expressive

What do you think about APAL versus PAL?

The easy direction. $PAL \subseteq APAL$: APAL subsumes PAL

The not so easy direction. $APAL \subseteq PAL$?

[!] φ is quite powerful as it quantifies over **formulas with all propositional variables** (even those not explicitly present in φ) and over **formulas of arbitrary finite modal depth**

APAL versus PAL

Theorem. PAL and EL are equally expressive

[!] φ is quite powerful as it quantifies over **formulas with all propositional variables** (even those not explicitly present in φ) and over **formulas of arbitrary finite modal depth**

Theorem. APAL is more expressive than PAL and EL

There are **no reduction axioms for APAL**, hence we have to find a proper axiomatisation...

Axiomatisation of APAL

Language of
APAL

$\mathcal{APAL} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi \mid [\varphi]\varphi \mid [!]\varphi$

Semantics

$M, s \vDash [!]\varphi$ iff $\forall \psi \in \mathcal{PAL} : M, s \vDash [\psi]\varphi$

Axioms of EL and PAL

$[!]\varphi \rightarrow [\psi]\varphi$ with $\psi \in \mathcal{PAL}$

From $\{\eta([\psi]\varphi) \mid \psi \in \mathcal{PAL}\}$
infer $\eta([!]\varphi)$

Infinite number of premises

$$\frac{\eta([\psi_1]\varphi) \ \eta([\psi_2]\varphi) \ \eta([\psi_3]\varphi) \ \dots}{\eta([!]\varphi)}$$

We call such a rule **infinitary**

Axiomatisation of APAL

Axioms of EL and PAL

$[!]\varphi \rightarrow [\psi]\varphi$ with $\psi \in \mathcal{PAL}$

From $\{\eta([\psi]\varphi) \mid \psi \in \mathcal{PAL}\}$
infer $\eta([!]\varphi)$

Theorem. There is a sound and complete infinitary axiomatisation of APAL

Open Problem. Is there a finitary axiomatisation of APAL?

Overview of APAL

Axioms of EL and PAL

$[!] \varphi \rightarrow [\psi] \varphi$ with $\psi \in \mathcal{PAL}$

From $\{\eta([\psi] \varphi) \mid \psi \in \mathcal{PAL}\}$
infer $\eta([!] \varphi)$

Infinite number of premises

Open Problem. Is there a finitary axiomatisation of APAL?

Theorem. APAL is more expressive than PAL

Theorem. APAL is sound and complete

Theorem. SAT-APAL is undecidable

Theorem. Complexity of MC-APAL is PSPACE-complete

Arbitrary AML

Language of
AAML

$$\mathcal{AAML} \ni \varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Box_a \varphi \mid [N_t]\varphi \mid [\otimes]\varphi$$

Semantics

$$M_s \models [\otimes]\varphi \text{ iff } \forall N_t : M_s \models [N_t]\varphi$$

$$M_s \models \langle \otimes \rangle \varphi \text{ iff } \exists N_t : M_s \models \langle N_t \rangle \varphi$$

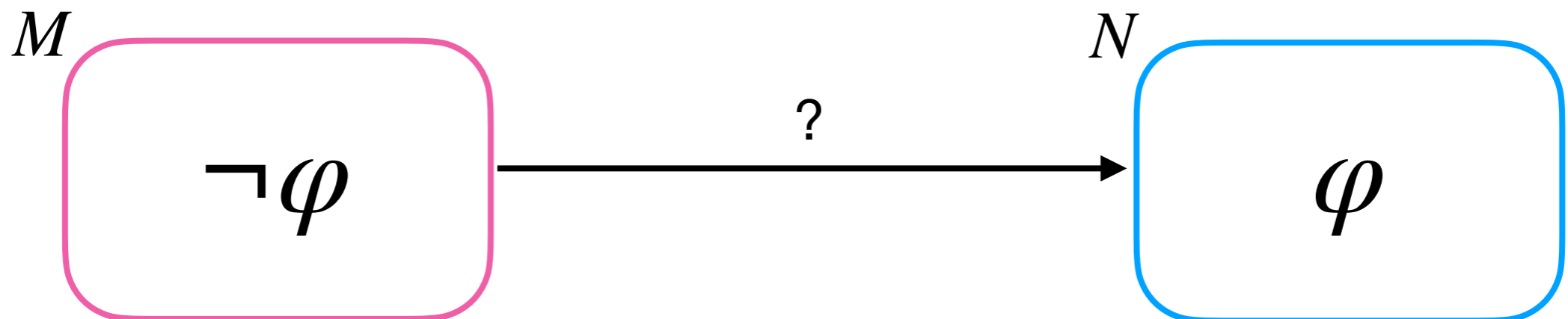
Preconditions are restricted to formulas without quantification

Synthesis

Synthesis Problem. Given a satisfiable formula φ ,
construct an action model N_X^φ such that

$$M_s \models \langle N_X^\varphi \rangle \varphi \text{ for any } M_s$$

Action models are so powerful that for a fixed goal we can construct one action model that will reach the goal in any situation (if the goal is reachable in principle)

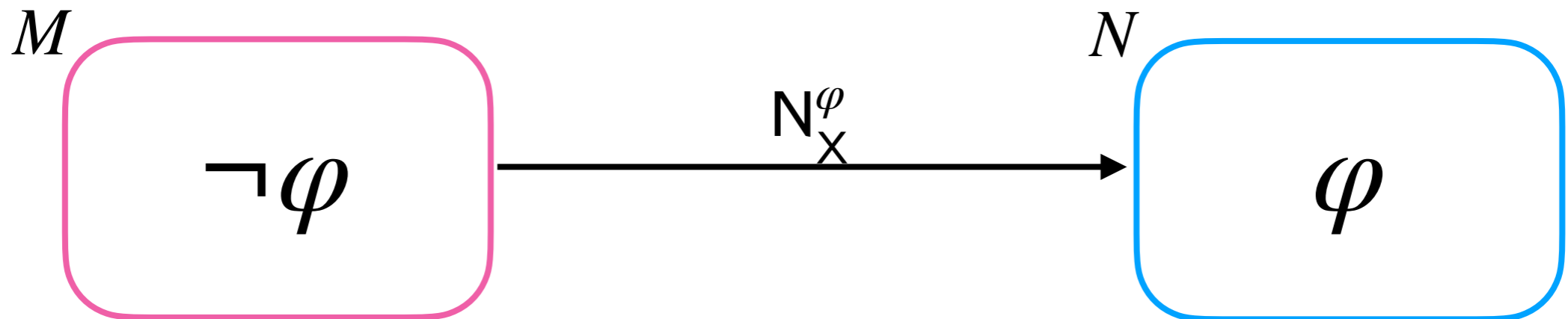


Synthesis

Synthesis Problem. Given a satisfiable formula φ ,
construct an action model N_X^φ such that

$$M_s \models \langle N_X^\varphi \rangle \varphi \text{ for any } M_s$$

Action models are so powerful that for a fixed goal we can construct one action model that will reach the goal in any situation (if the goal is reachable in principle)



Synthesis

Synthesis Problem. Given a satisfiable formula φ ,
construct an action model N_X^φ such that
$$M_s \models \langle N_X^\varphi \rangle \varphi \text{ for any } M_s$$

Synthesis of such action models is **possible**

But what is the connection between the **synthesis problem** and **quantification** over action models?

Synthesis Problem*. Given a formula φ , construct an
action model N_X^φ such that $\models \langle \otimes \rangle \varphi \leftrightarrow \langle N_X^\varphi \rangle \varphi$

Disclaimer: we assume that \sim in M is arbitrary

Synthesis

Synthesis Problem*. Given a formula φ , construct an action model N_X^φ such that $\models \langle \otimes \rangle \varphi \leftrightarrow \langle N_X^\varphi \rangle \varphi$

Wait! What???

Schema $\langle \otimes \rangle \varphi \leftrightarrow \langle N_X^\varphi \rangle \varphi$ is a **reduction axiom** for AAML

This implies something crazy...

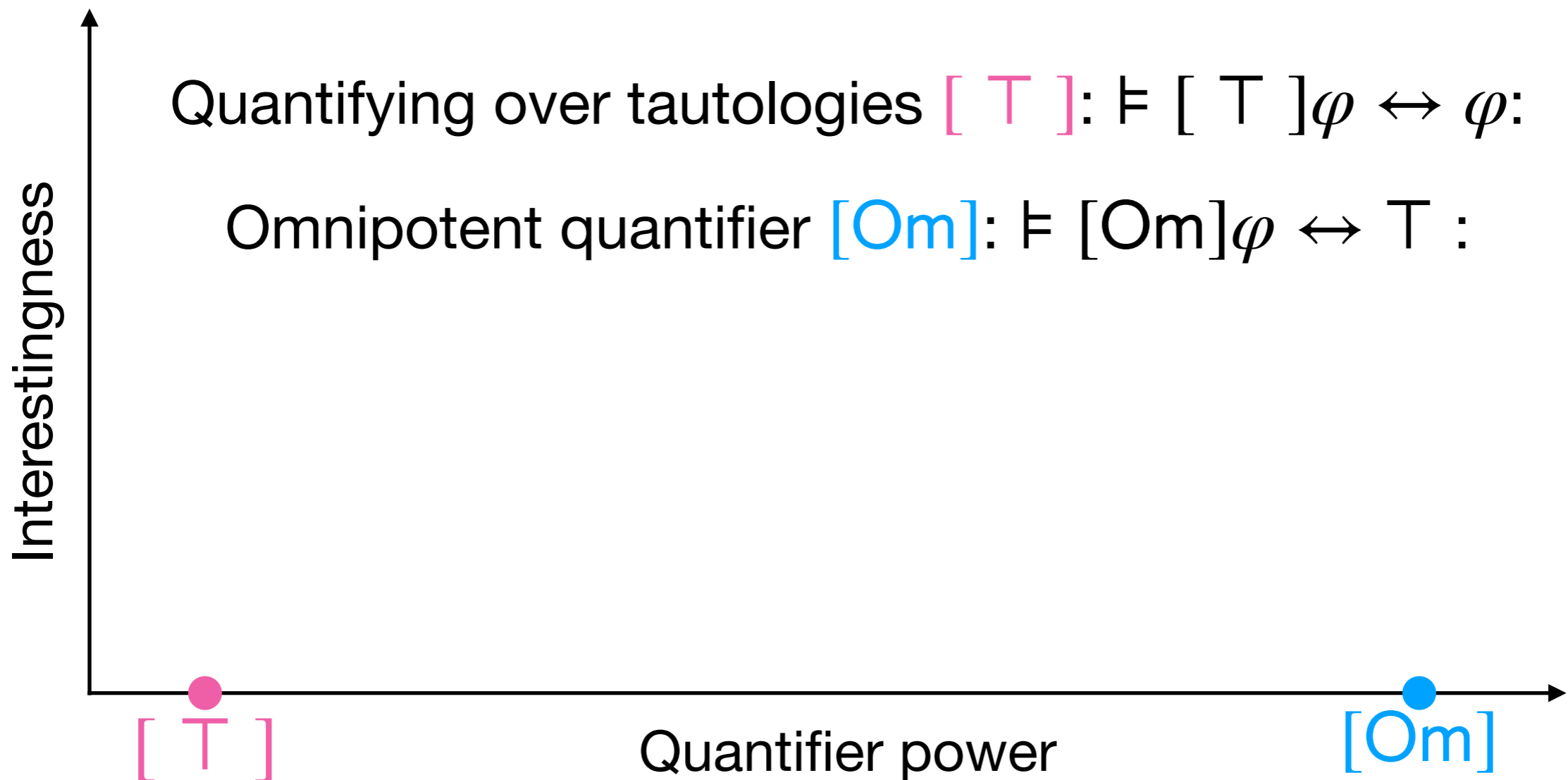
Theorem. AAML is as expressive as EL

Theorem. APAL is more expressive than PAL and EL

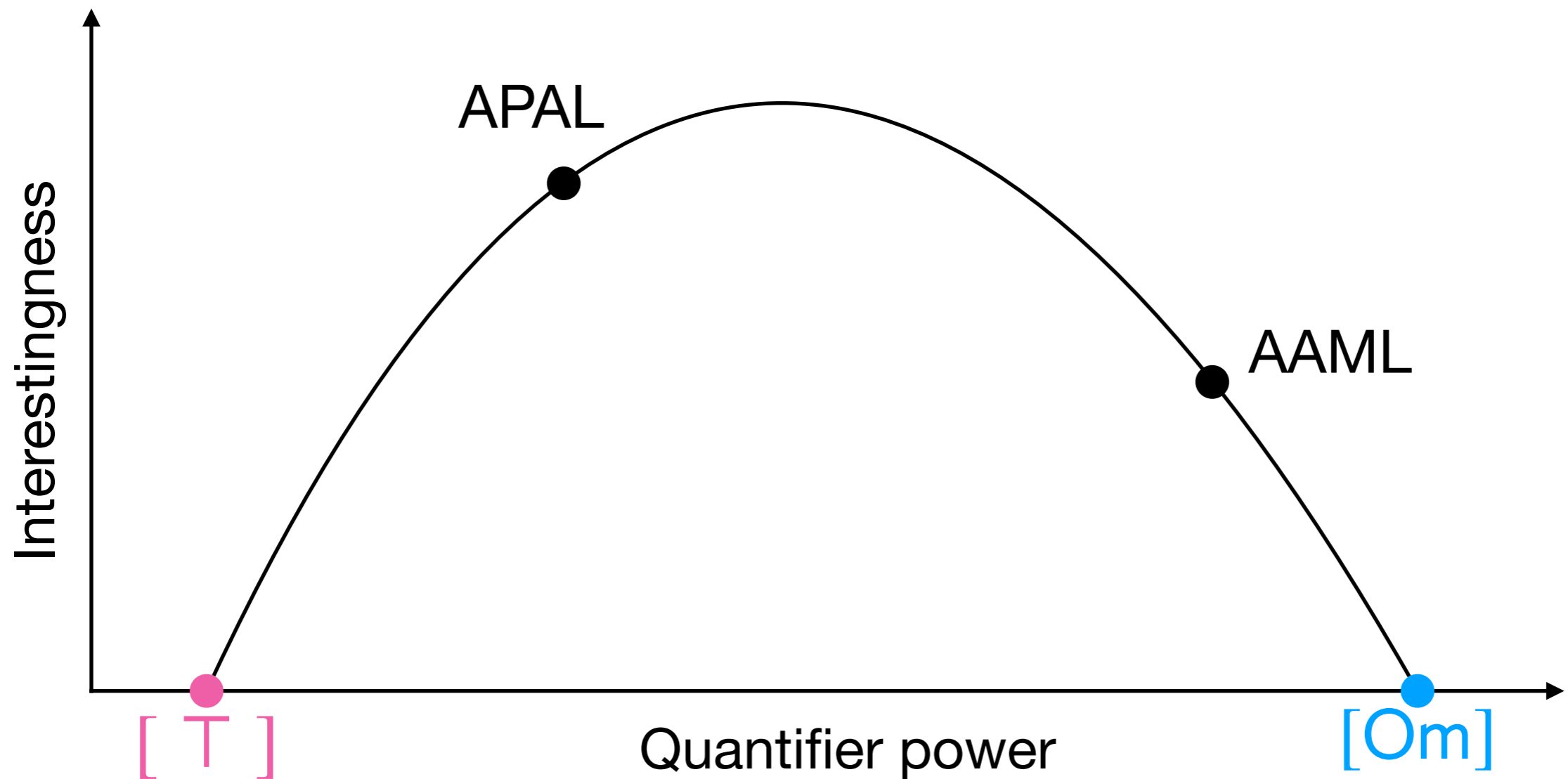
Theorem. AAML is decidable

Theorem. APAL is undecidable

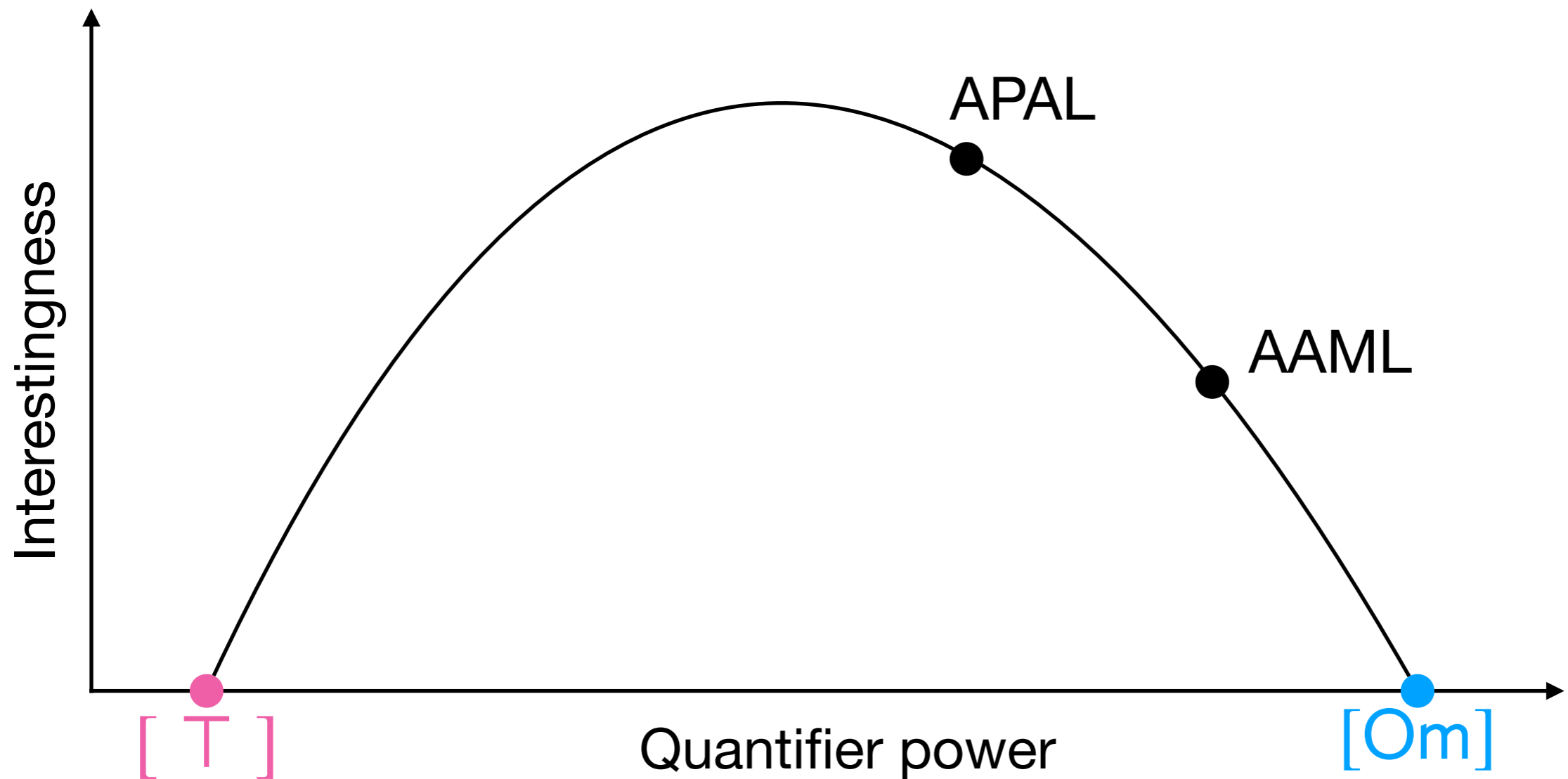
Interestingness of Quantification



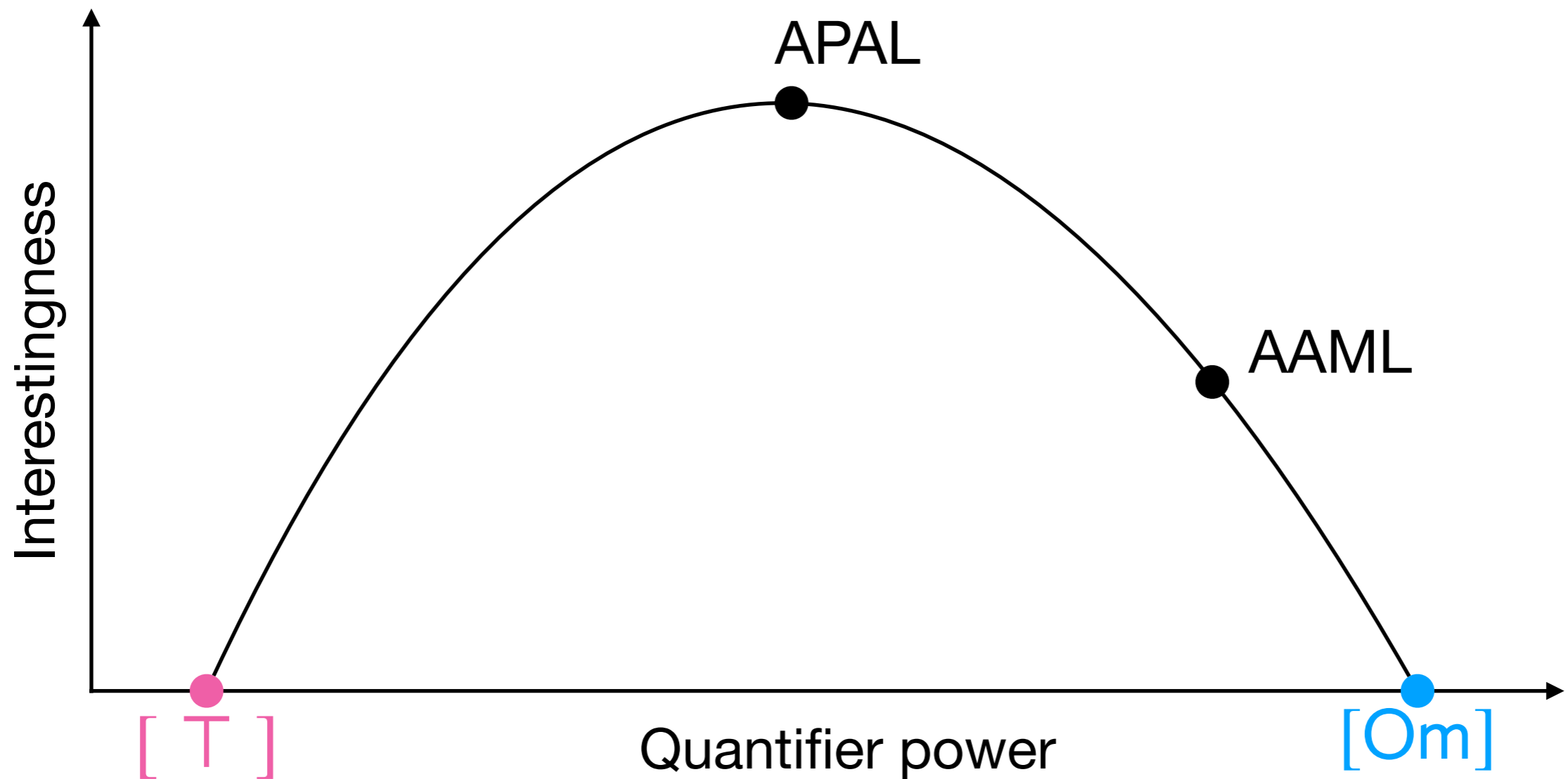
Interestingness of Quantification



Interestingness of Quantification



Interestingness of Quantification



Quantification Overview

- **Shifts the emphasis** from particular epistemic updates to (non-)existence of an update reaching a certain goal
- **Fun and unpredictable**: APAL is highly complex, while AAML is technically the same as EL
- A **powerful tool** for DEL-inspired logics. E.g. existence of a posting strategy in social network logics, etc.
- Lots of tantalising **open questions!**

Open Problem. Is there a finitary axiomatisation of APAL?

If You Want More

Information and Computation 292 (2023) 105026



Contents lists available at [ScienceDirect](#)

Information and Computation

journal homepage: www.elsevier.com/locate/yinco



To be announced

Hans van Ditmarsch

University of Toulouse, CNRS, IRIT, France



ARTICLE INFO

Article history:

Received 1 July 2021
Received in revised form 14 March 2023
Accepted 15 March 2023
Available online 21 March 2023

ABSTRACT

In this survey we review dynamic epistemic logics with modalities for quantification over information change. Of such logics we present complete axiomatizations, focussing on axioms involving the interaction between knowledge and such quantifiers, we report on their relative expressivity, on decidability and on the complexity of model checking and satisfiability, and on applications. We focus on open problems and new directions for research.

© 2023 Elsevier Inc. All rights reserved.

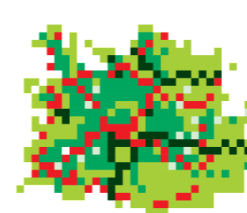
Quantification in Dynamic Epistemic Logic

Area: Logic and Computation (LoCo)

Level: Introductory

Website: <https://rgalimullin.gitlab.io/esslli23.html>

Lecturer(s): Rustam Galimullin and Louwe B. Kuijer



ESLLI 2023

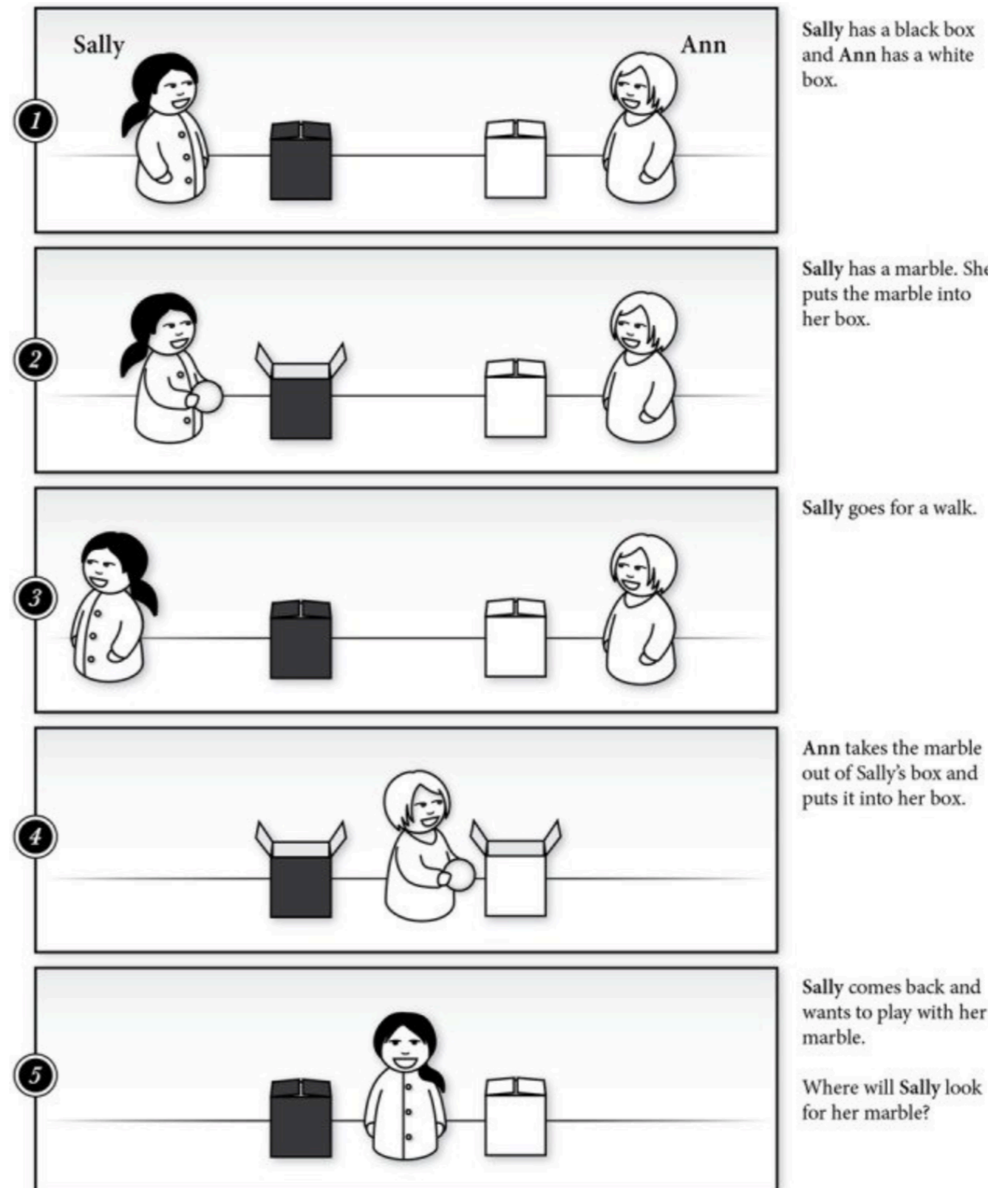
> LJUBLJANA > SLOVENIA

Theory of Mind

- In epistemic logic agents may have knowledge of not only their own knowledge but knowledge of others as well
- In other words, agents may have mental models of what other agents (mistakenly) 'think'
- Bob knows that the cat is in the house, and he also knows that Alice considers it possible that the cat is out
- Such a capacity to ascribe mental states to other agents is called theory of mind

Sally-Anne Test

- Ability of human (and artificial) agents to ascribe false beliefs to other agents may be checked by the Sally-Anne test
- The test was developed in 1985 by psychologists researching cognitive abilities of children

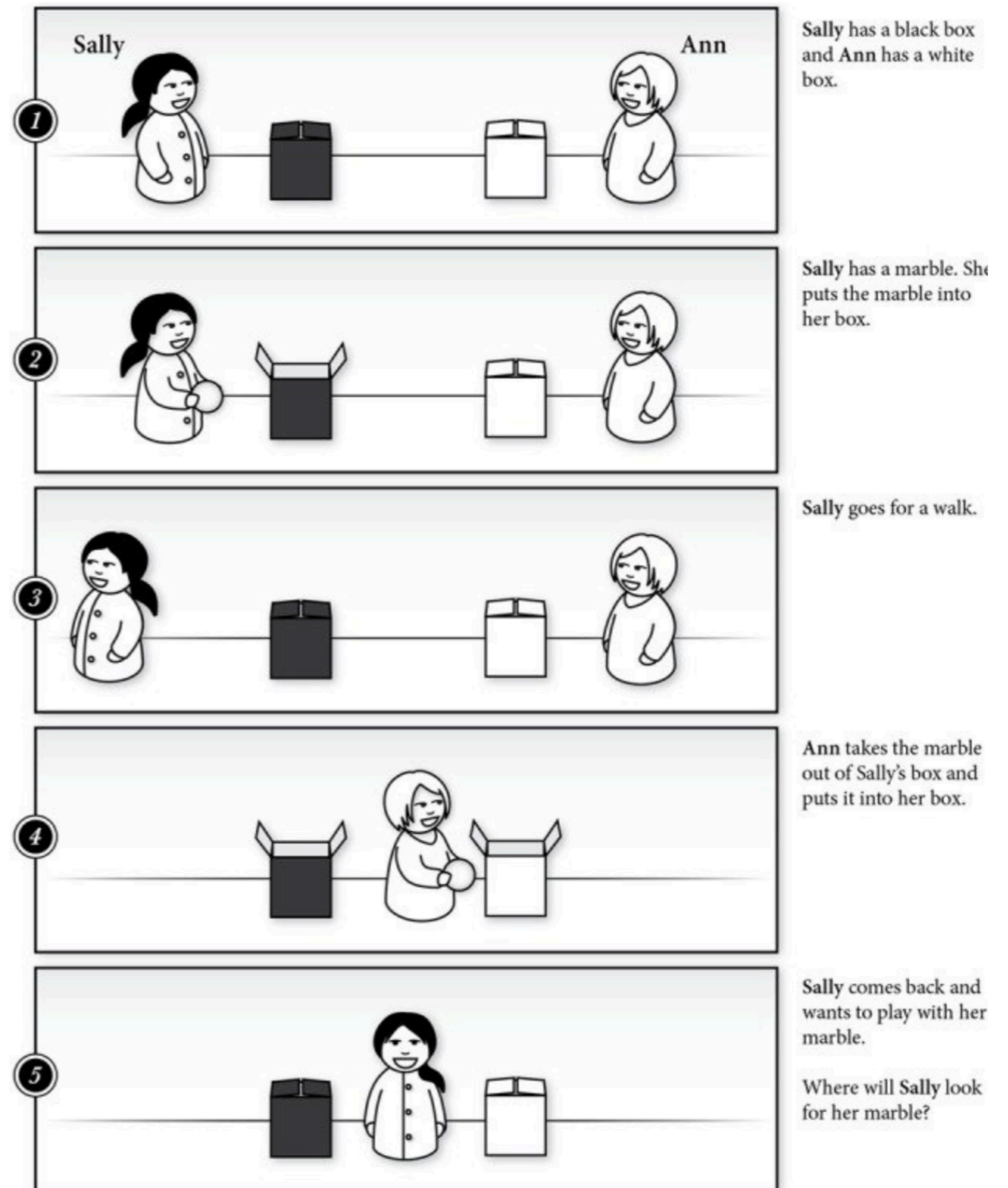


Sally-Anne Test in DEL

Before we formalise the test in DEL, look at the figure and think why **action models do not quite work here...**

First, we need to be able to **change basic facts of the world** (e.g. marble being transferred from one box to another)

Second, we need to be able to **reason about (false) beliefs**, rather than knowledge

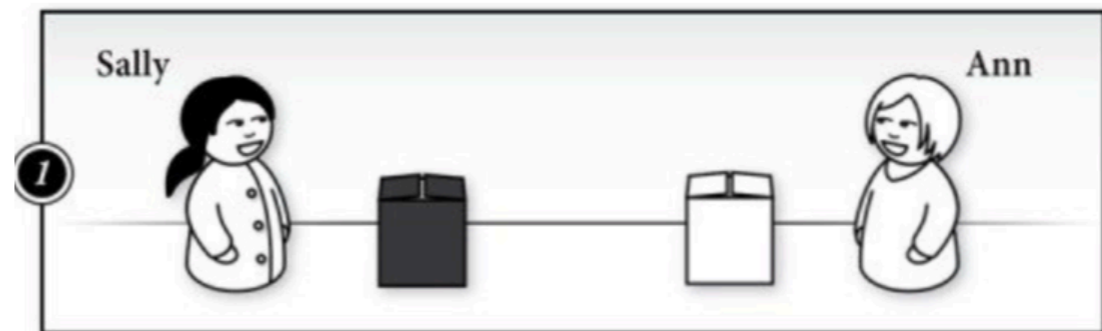
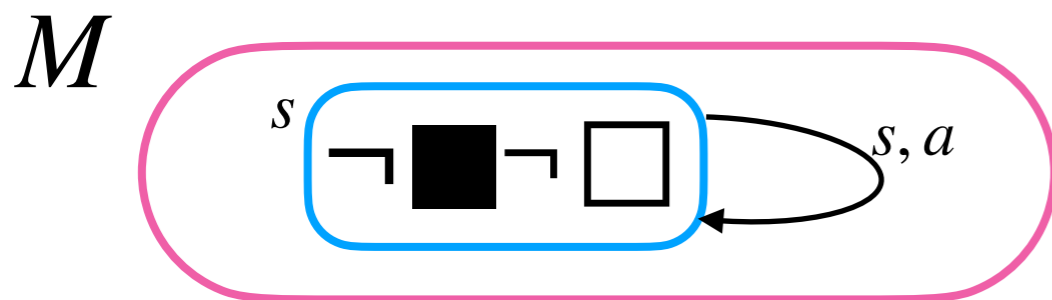


Sally-Anne Test in DEL

Epistemic models

An **epistemic model** M is a tuple (S, \sim, V) , where

- $S \neq \emptyset$ is a set of states;
- $\sim: A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an arbitrary relation;
- $V: P \rightarrow 2^S$ is the valuation function.

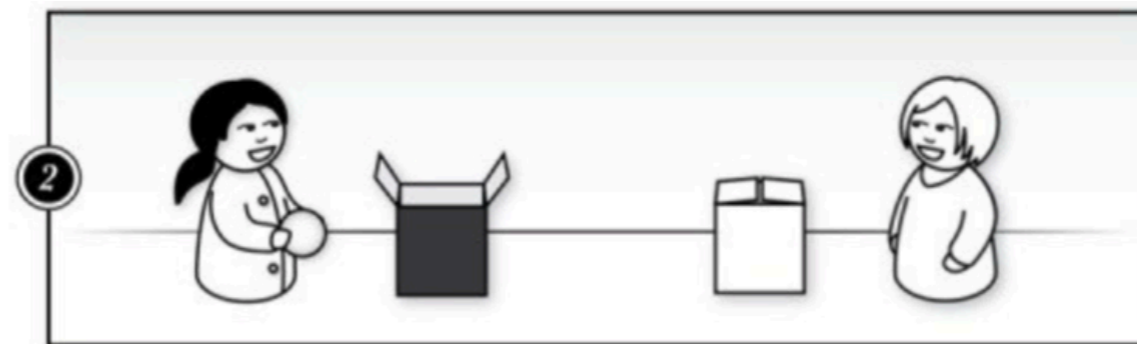
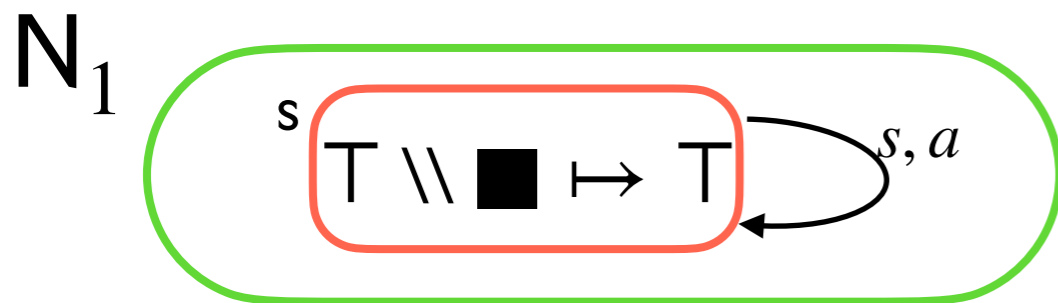


■ (□): the marble is in the black (white) box

Sally-Anne Test in DEL

An **action model** N is a tuple (S, \sim, pre) , where

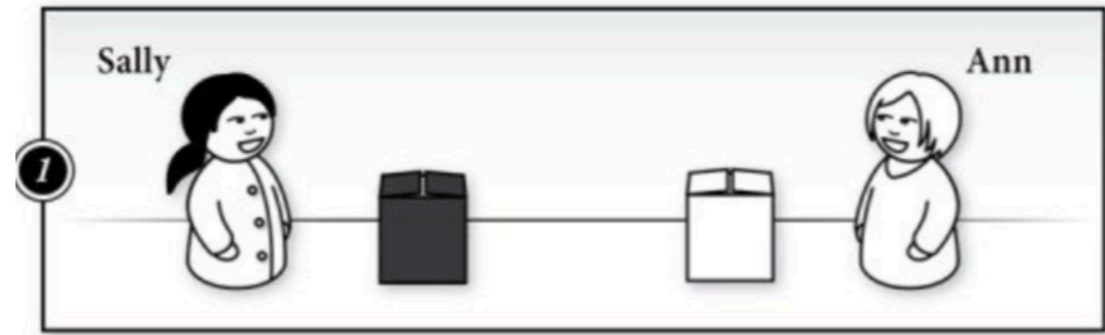
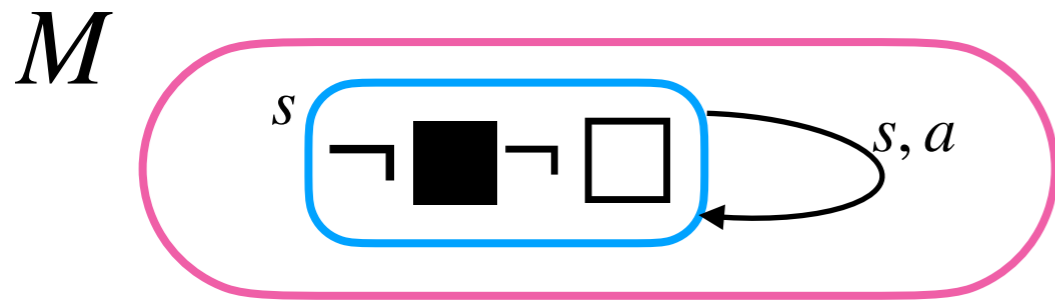
- $S \neq \emptyset$ is a set of states;
- $R : A \rightarrow 2^{S \times S}$ is an indistinguishability function with each \sim_a being an **arbitrary relation**;
- $\text{pre} : S \rightarrow \mathcal{L}$ is the precondition function;
- $\text{post} : S \rightarrow (P \rightarrow \mathcal{L})$ is the **postcondition function**, assigning in each state postconditions for finitely many propositional variables.



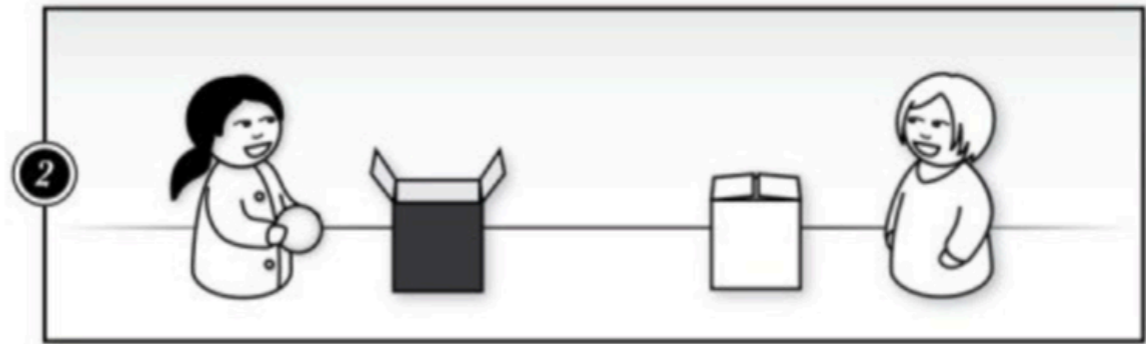
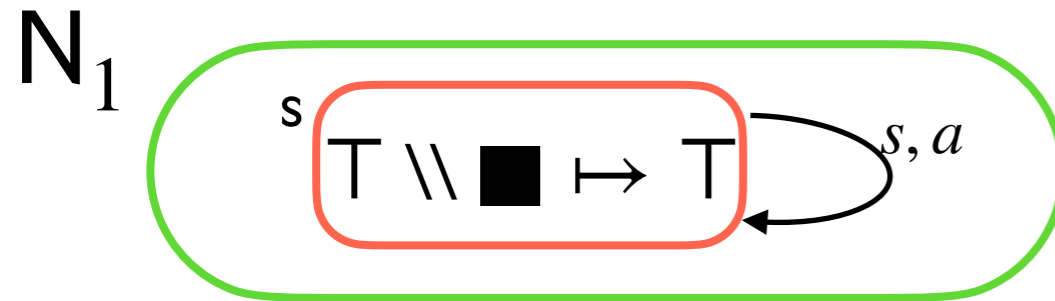
Sally has a marble. She puts the marble into her box.

■ (□): the marble is in the black (white) box

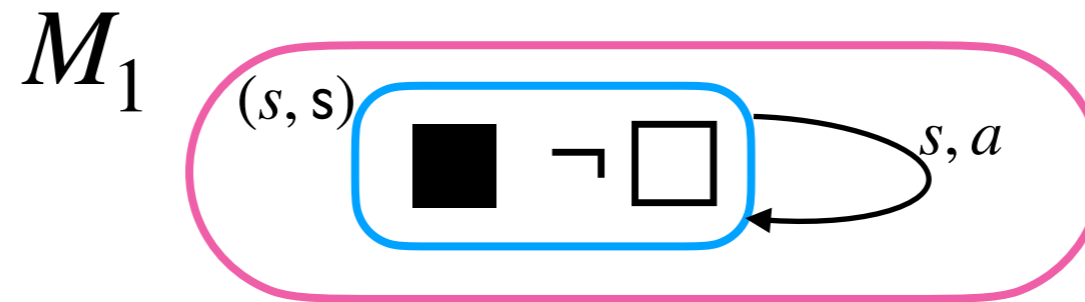
Sally-Anne Test in DEL



Sally has a black box and Ann has a white box.

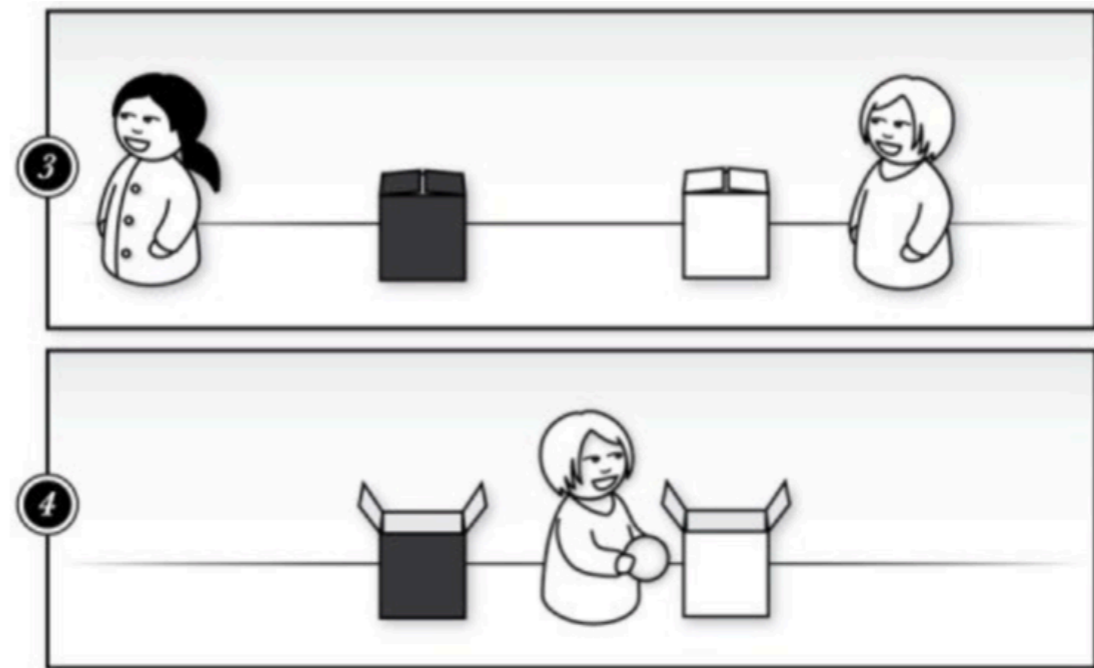
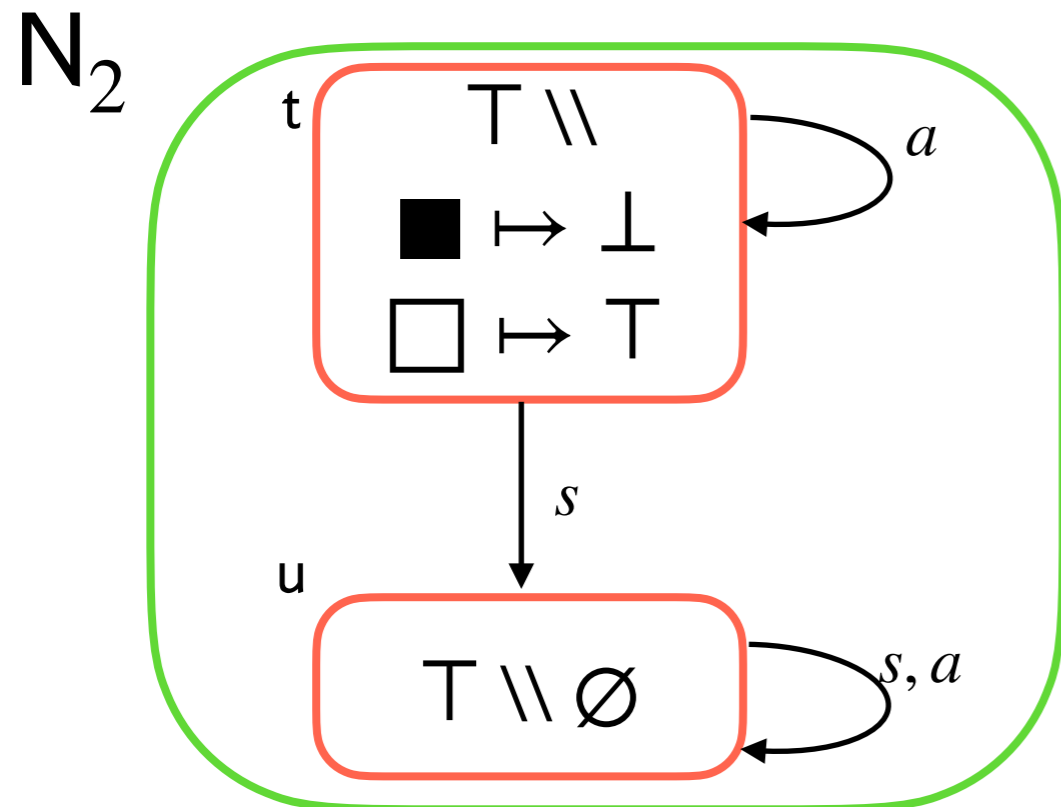
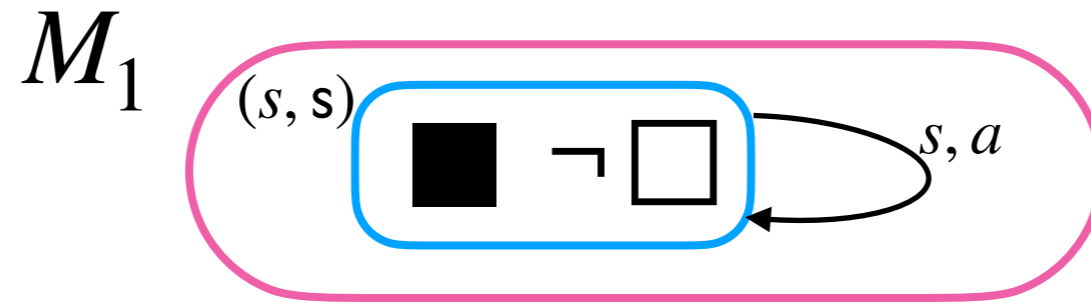


Sally has a marble. She puts the marble into her box.



■ (□): the marble is in the black (white) box

Sally-Anne Test in DEL

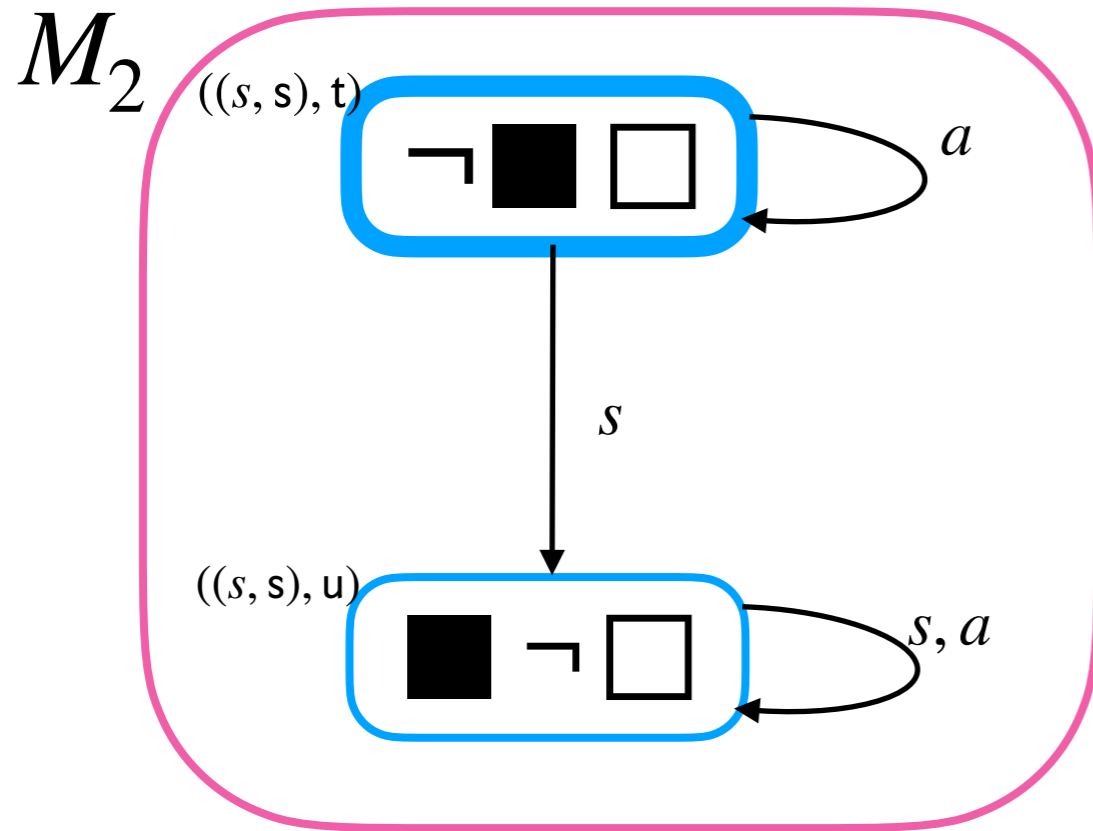


Sally goes for a walk.

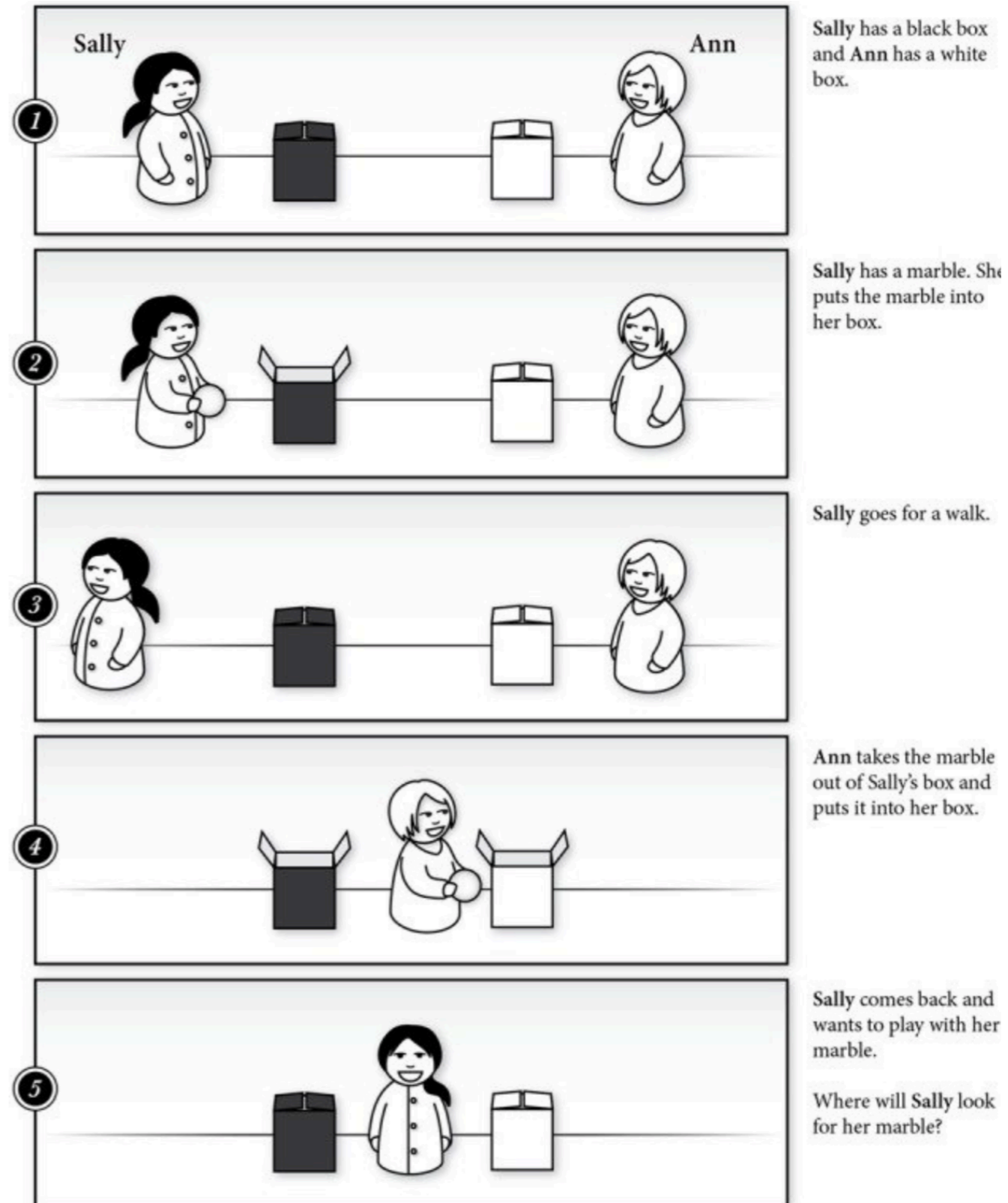
Ann takes the marble out of Sally's box and puts it into her box.

\blacksquare (\square): the marble is in the black (white) box

Sally-Anne Test in DEL



Anne knows the state of affairs, while Sally believes that the marble is in the black box (while it is actually in the white one)



Social Robotics

- While modelling theory of mind and false-belief tasks in DEL is interesting in itself, it has some interesting **prospective applications to multi-agent systems**
- Interaction of human and artificial agents calls for **socially-aware robotics**
- <https://www.ijcai.org/proceedings/2020/224>

Where to Start

Seeing is Believing: Formalising False-Belief Tasks in Dynamic Epistemic Logic

Thomas Bolander

Technical University of Denmark

Implementing Theory of Mind on a Robot Using Dynamic Epistemic Logic

Lasse Dissing, Thomas Bolander

 [Short video](#)

 [Long video](#)

Thank you!