True Lies

Xixi Logic Seminar Series
Zhejiang University, 10 December 2014

Thomas Ågotnes
University of Bergen, Norway

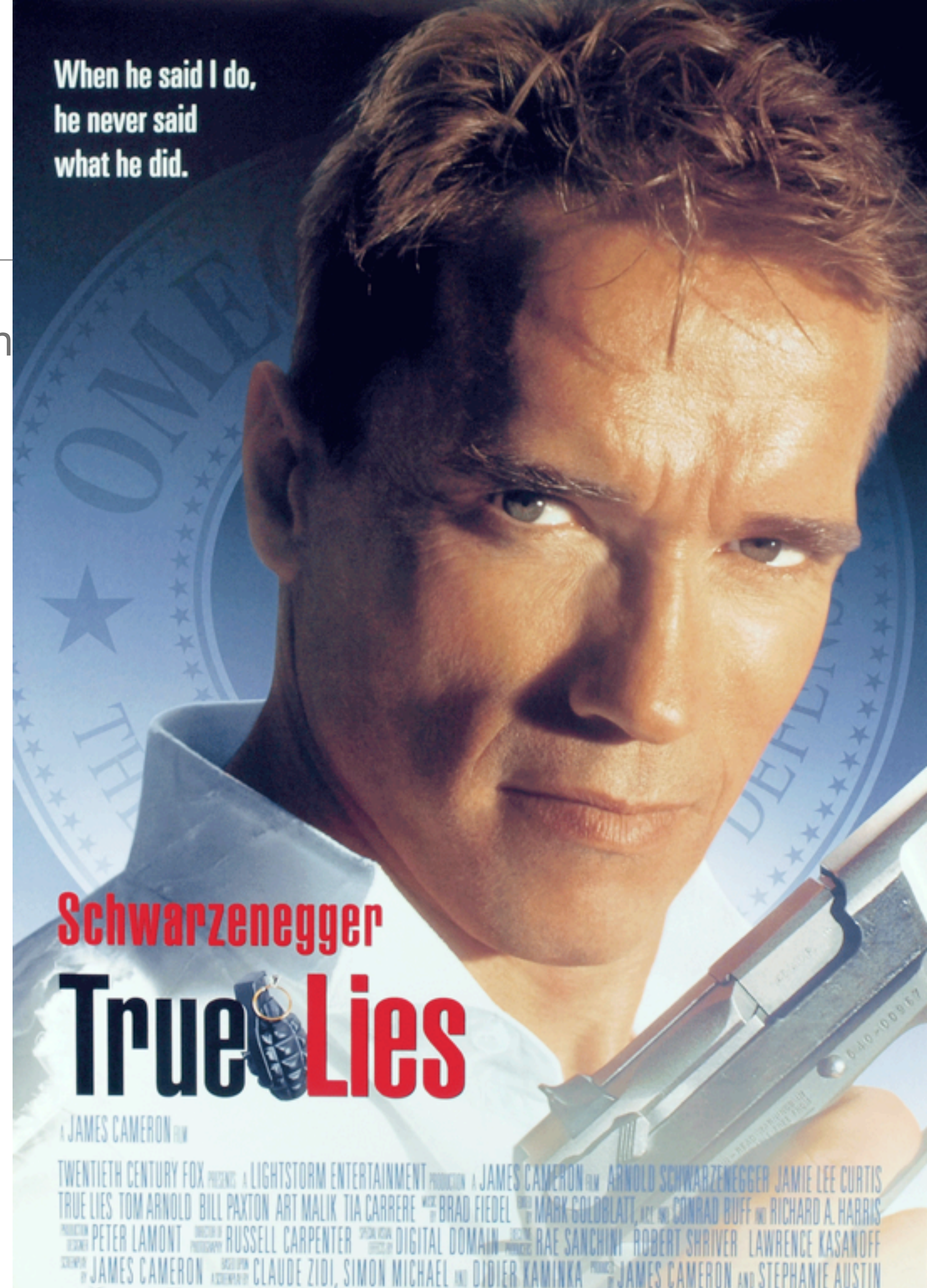
Joint work with
Hans van Ditmarsch
Yanjing Wang

Introduction

- A true (or self-fulfilling) lie, is a lie that becomes true when it is made
- Example: Thomas' party
- Logical vs. non-logical true lies
- Outline:
  - Background
  - Public true lies
  - Private true lies

Introduction

- A true (or self-fulfilling) lie, is a lie th
- Example: Thomas' party
- Logical vs. non-logical true lies
- Outline:
  - Background
  - Public true lies
  - Private true lies



Introduction and Background

Modal logics of knowledge and belief

$\varphi ::= p \mid B_i\varphi \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2$

Dual: $\hat{B}_i\phi \equiv \neg B_i \neg\phi$

$$M = (S, \sim_1, \ldots, \sim_n, V) \qquad \sim_i \text{ accessibility rel. over S}$$

$$M, s \models B_i\phi \qquad \Leftrightarrow \qquad \forall t \sim_i s \ M, t \models \phi$$

Modal logics of knowledge and belief

$\varphi ::= p \mid B_i\varphi \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2$ Dual: $\hat{B}_i\phi \equiv \neg B_i \neg\phi$

$M = (S, \sim_1, \ldots, \sim_n, V)$ $\sim_i$ accessibility rel. over S

$$M, s \models B_i\phi \quad \Leftrightarrow \quad \forall t \sim_i s \; M, t \models \phi$$

If we want to model *knowledge* rather than *belief* we assume that each $\sim_i$ is a equivalence relation.

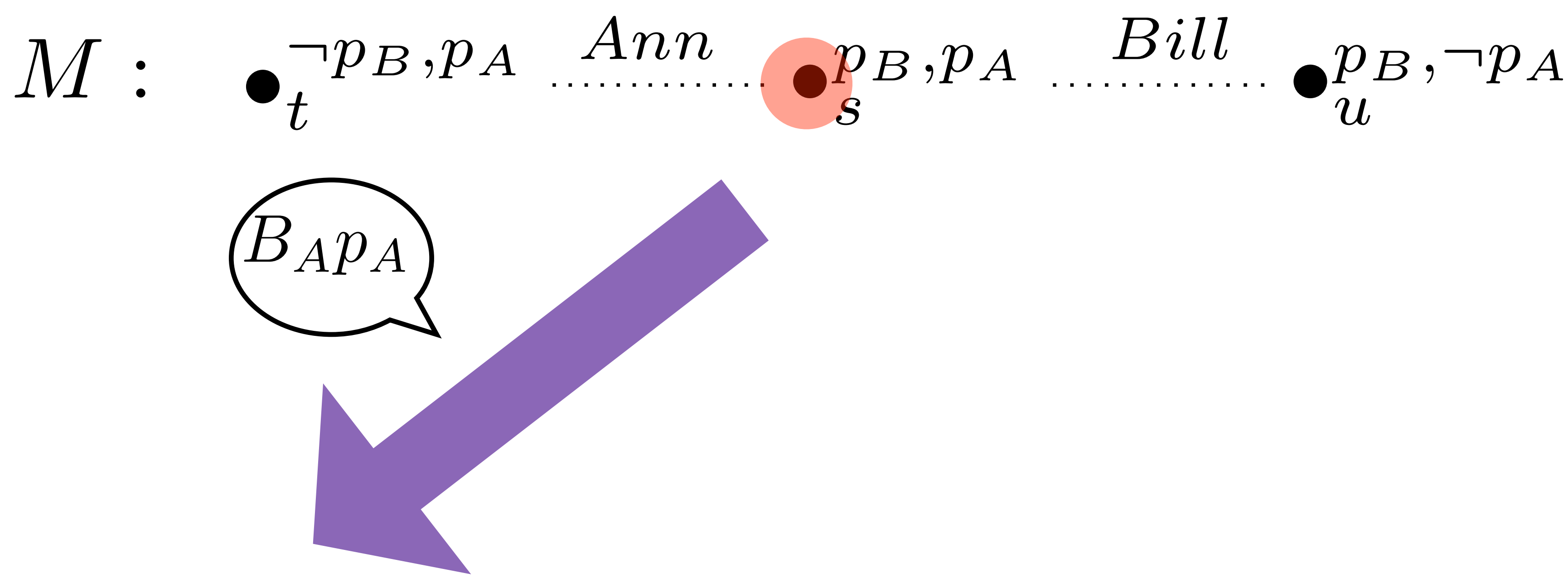
Learning by removing possibilities (knowledge)

M :

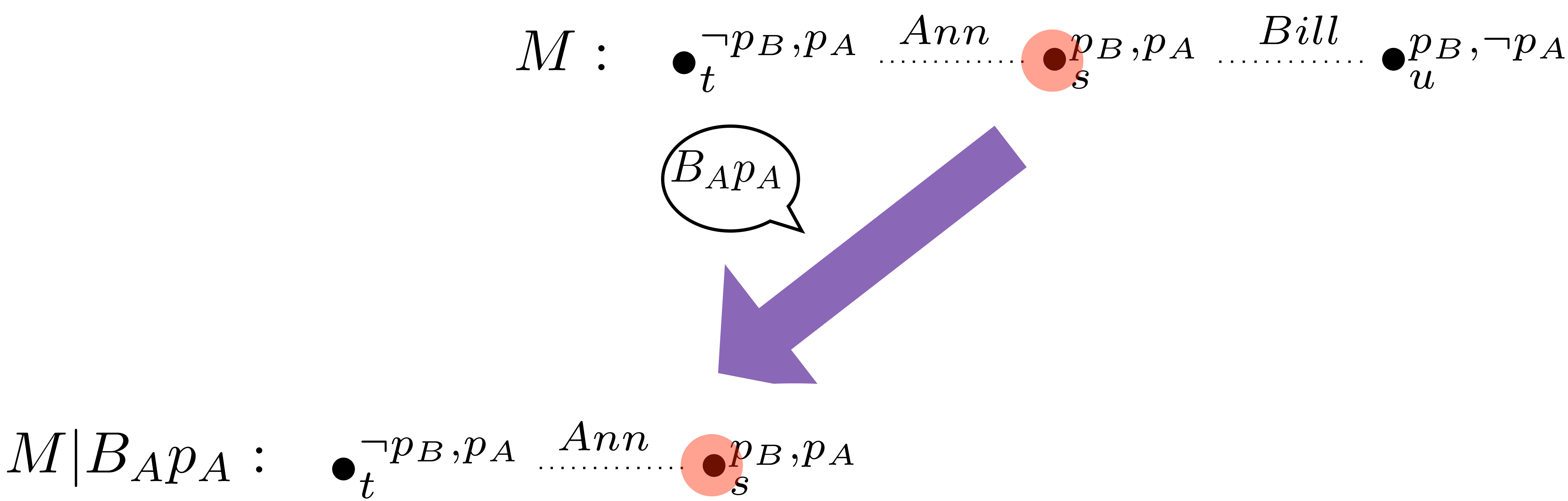
Learning by removing possibilities (knowledge)

$$M: \quad \bullet_{t}^{\neg p_B, p_A} \;\overset{Ann}{\cdots\cdots\cdots}\; \bullet_{s}^{p_B, p_A} \;\overset{Bill}{\cdots\cdots\cdots}\; \bullet_{u}^{p_B, \neg p_A}$$

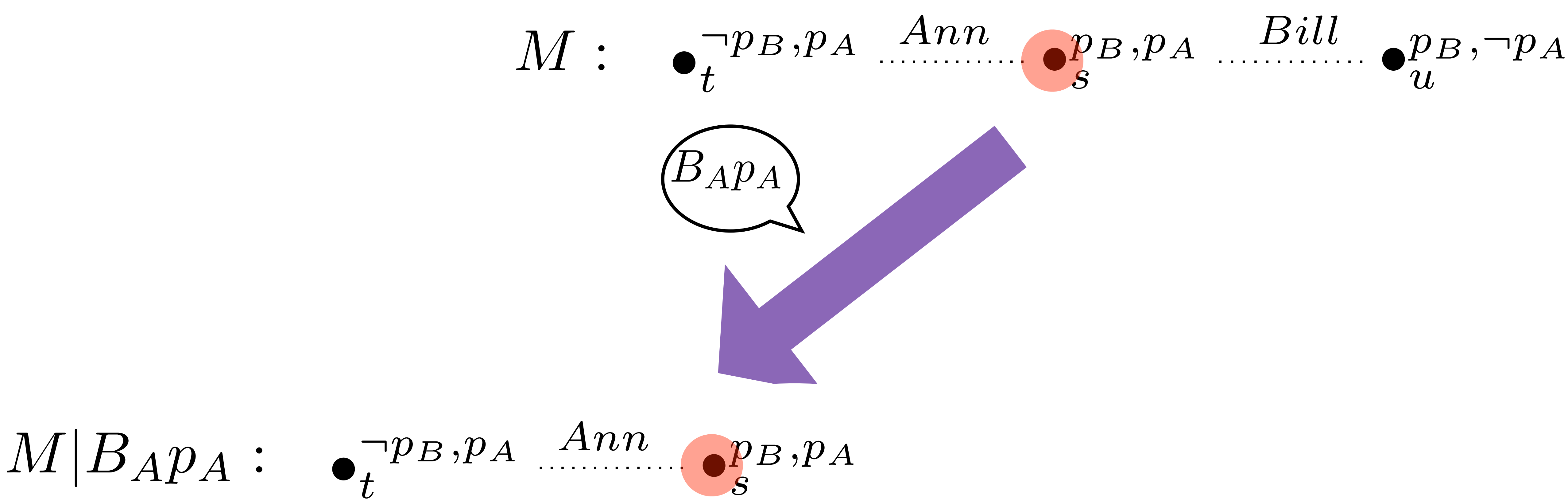
Learning by removing possibilities (knowledge)



Learning by removing possibilities (knowledge)

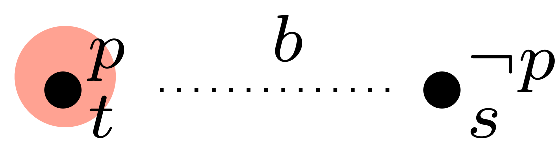


Learning by removing possibilities (knowledge)



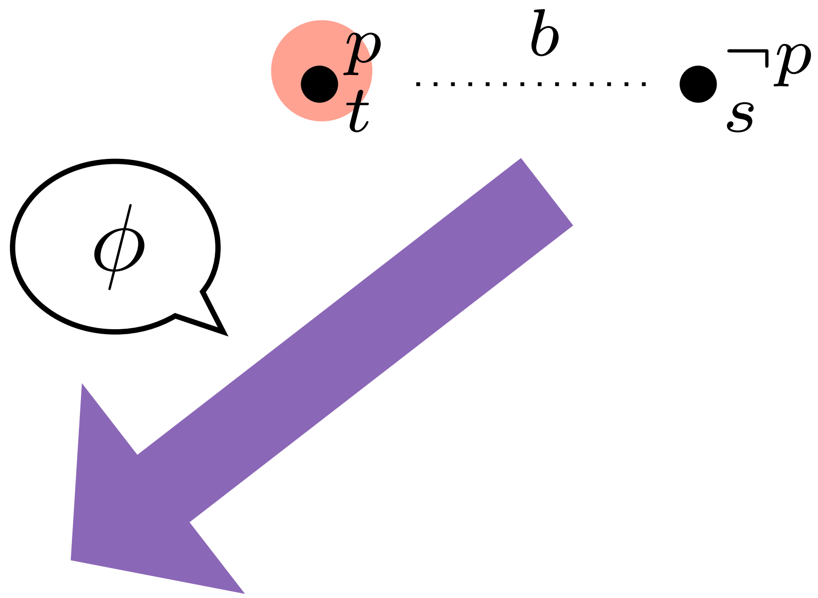
$$M|B_A p_A, s \models B_B B_A p_A$$

Moore sentences



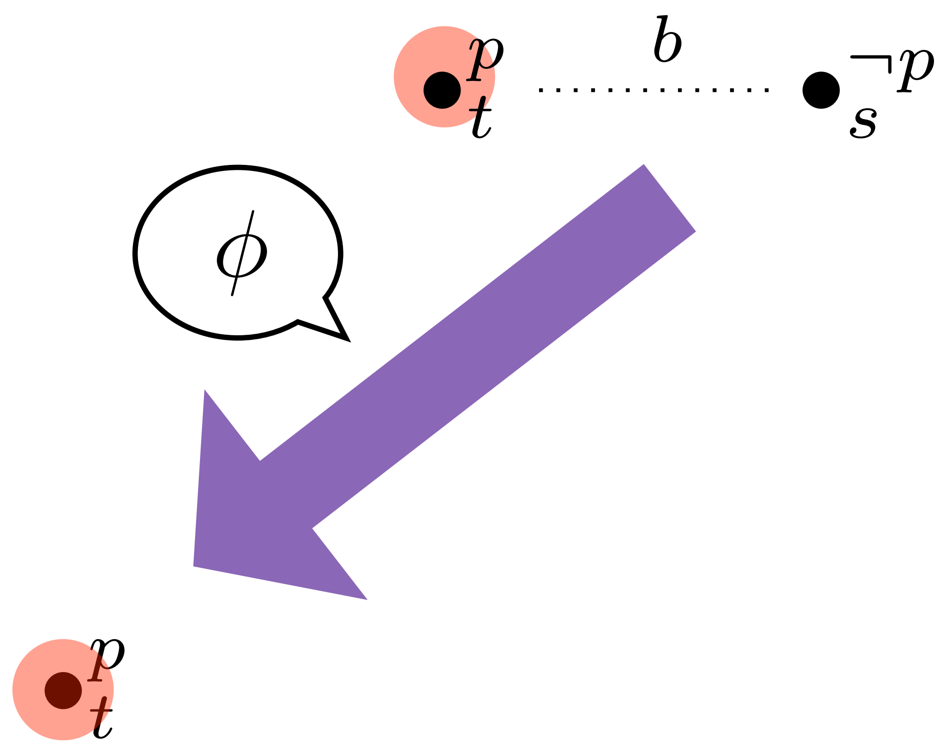
$$\phi = p \wedge \neg B_b p$$

Moore sentences



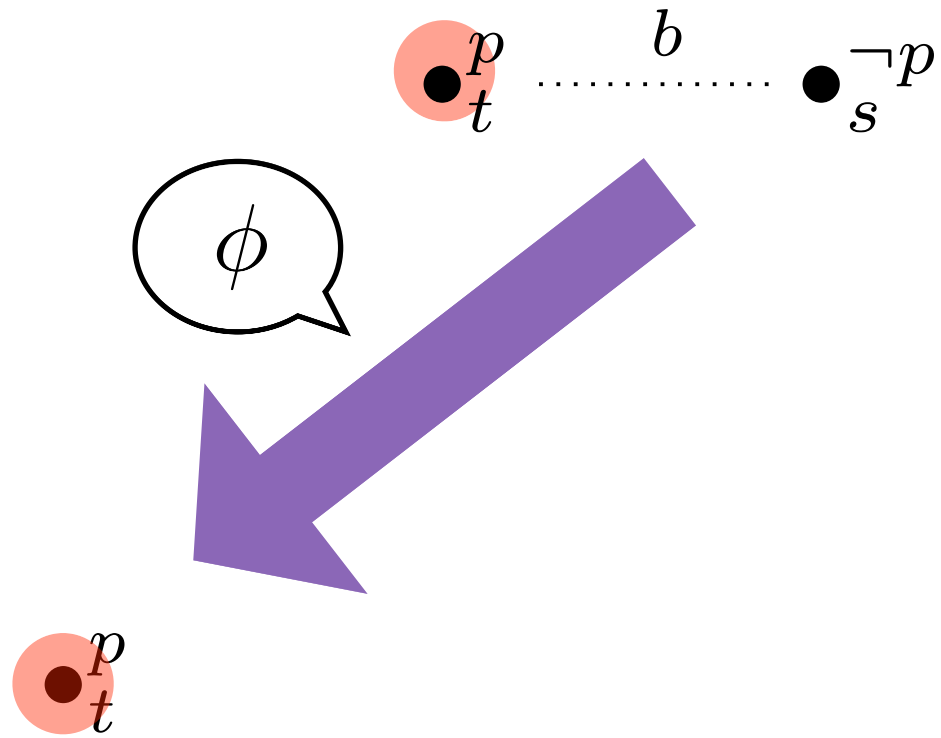
$$\phi = p \wedge \neg B_b p$$

Moore sentences



$$\phi = p \wedge \neg B_b p$$

Moore sentences



$$\phi = p \land \neg B_b p$$

$$M|_\phi, s \models \neg\phi$$

Lies

- Dimensions:
  - Who is the lier: one of the agents in the system, or an outsider?
  - Who are being lied to (and what do the others know about that)?
  - What are the agents’ attitudes to possible lies?
    - Credulous agents: believe everything
    - Skeptical agents: believe everything consistent with their existing beliefs
    - ...

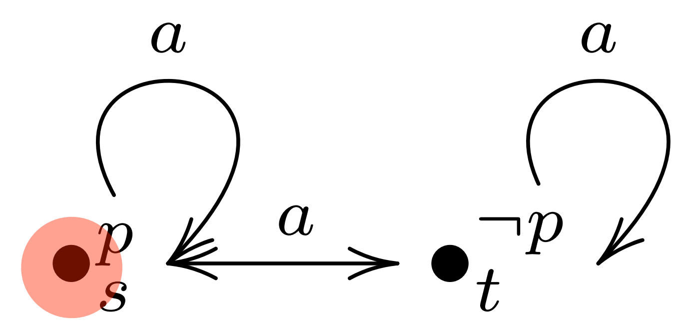
Lies

- Here:
  - Two cases: one of the agents in the system + outside observer
  - Credulous/skeptical agents
  - Public lie, to all other agents
  - Private lies

Public true lies from the outside

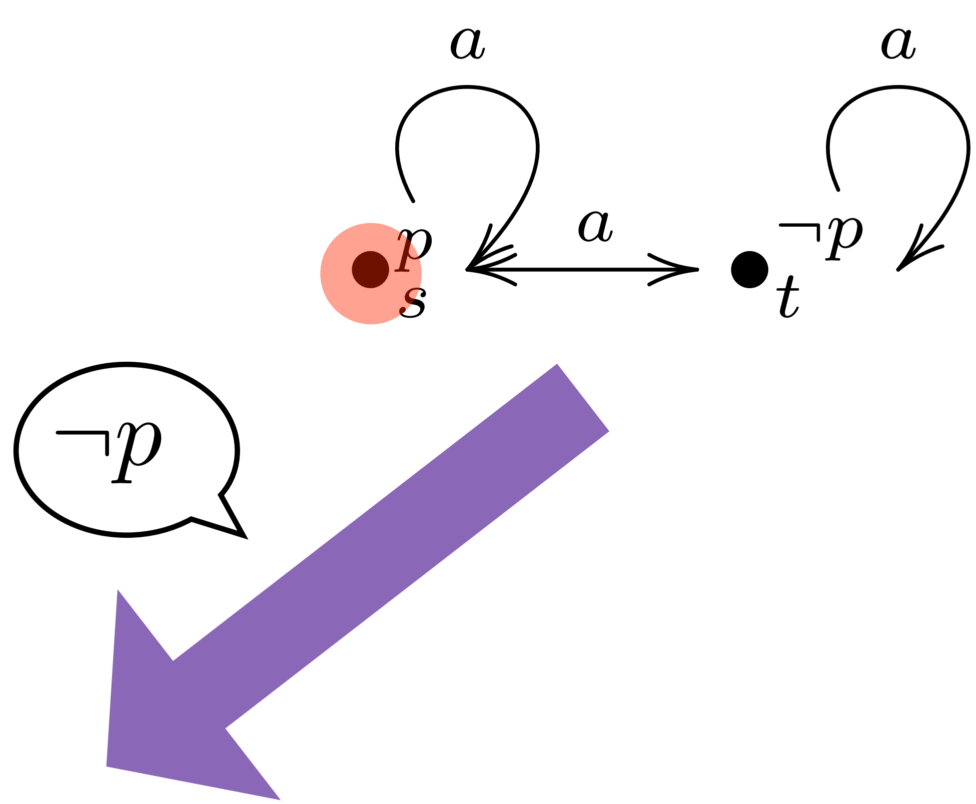
Untruthful announcements: link-cutting semantics

$M:$

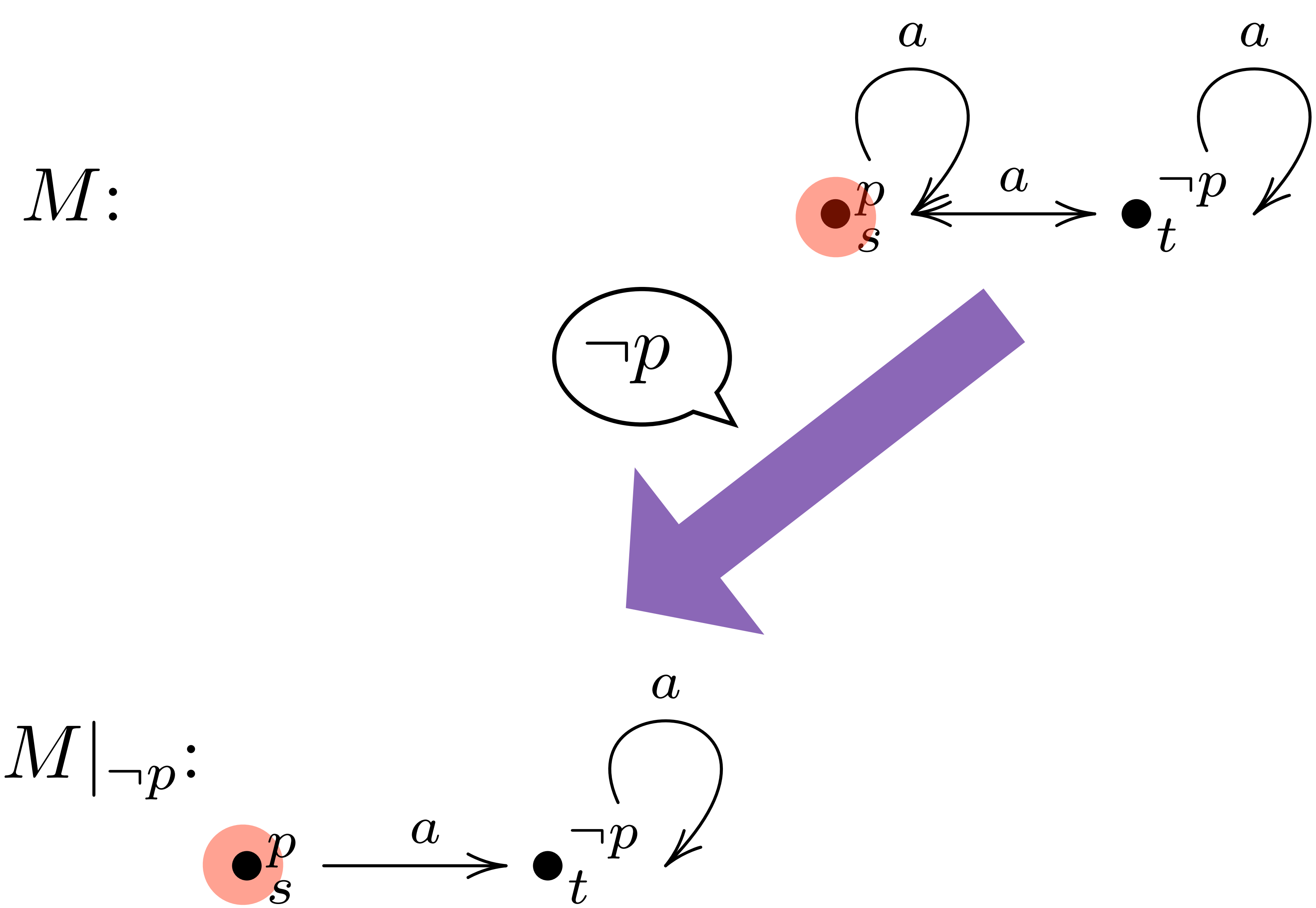


Untruthful announcements: link-cutting semantics

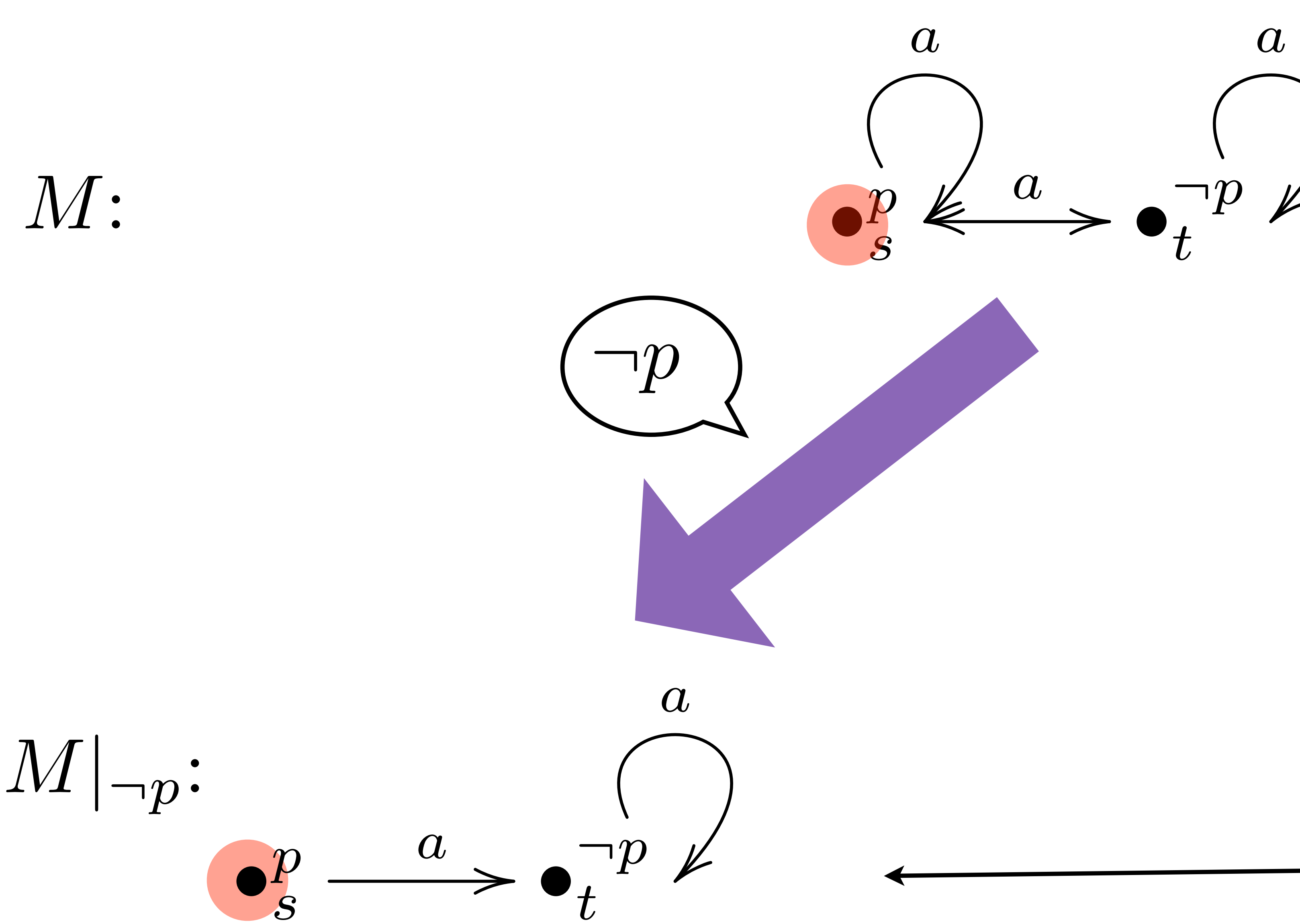
M:



Untruthful announcements: link-cutting semantics



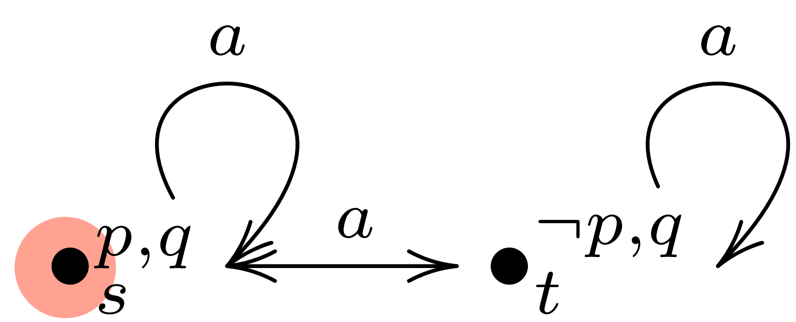
Untruthful announcements: link-cutting semantics



Update obtained by removing links going into states where the announcement is false

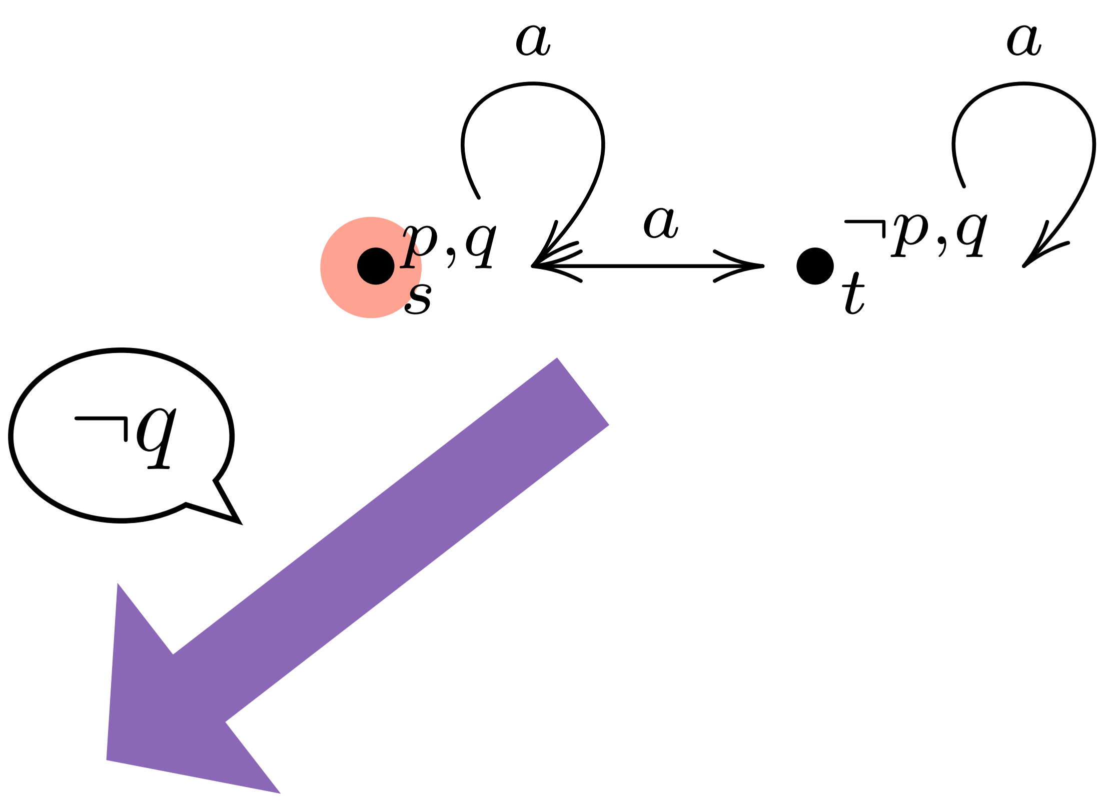
Unbelievable lie

M:

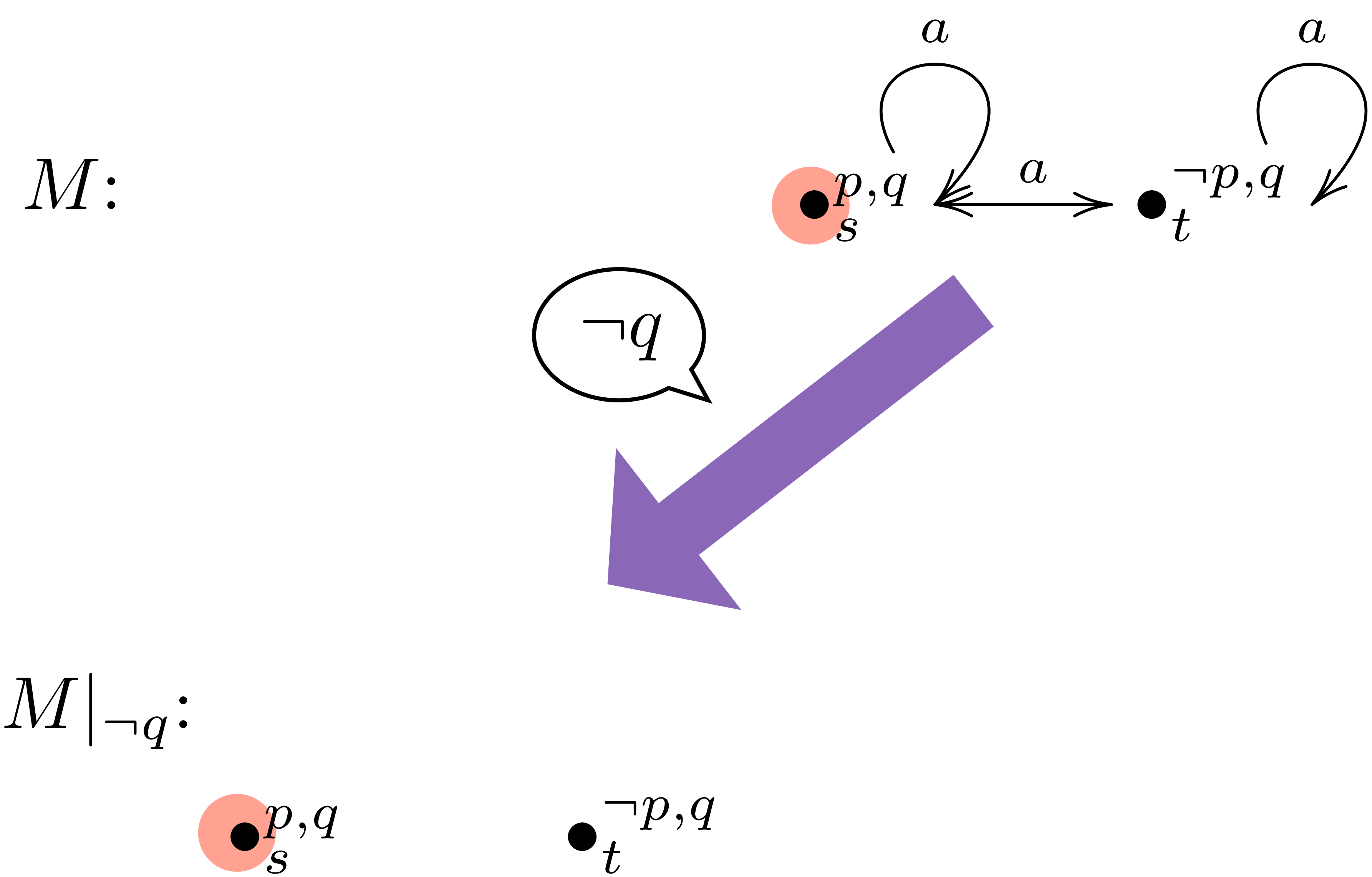


Unbelievable lie

M:

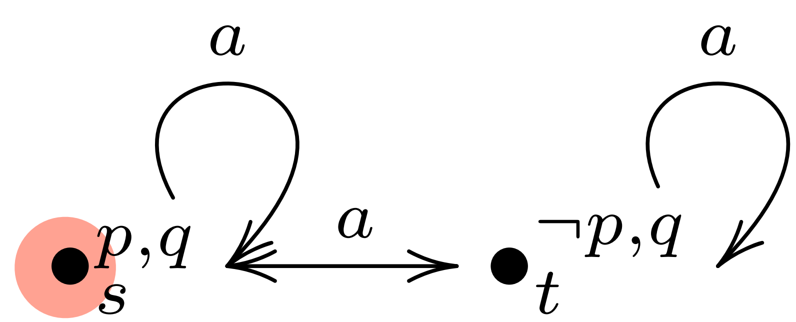


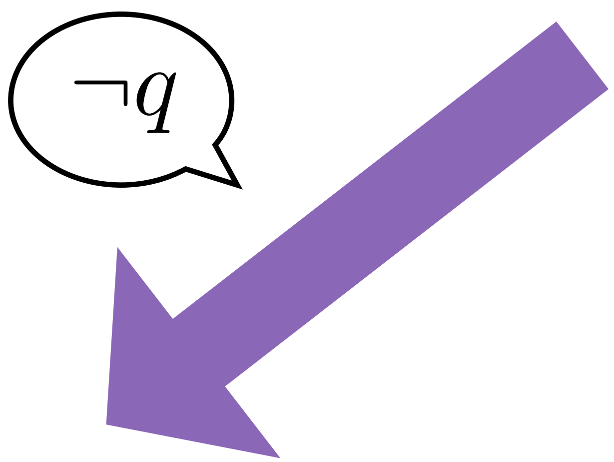
Unbelievable lie



Unbelievable lie

M:

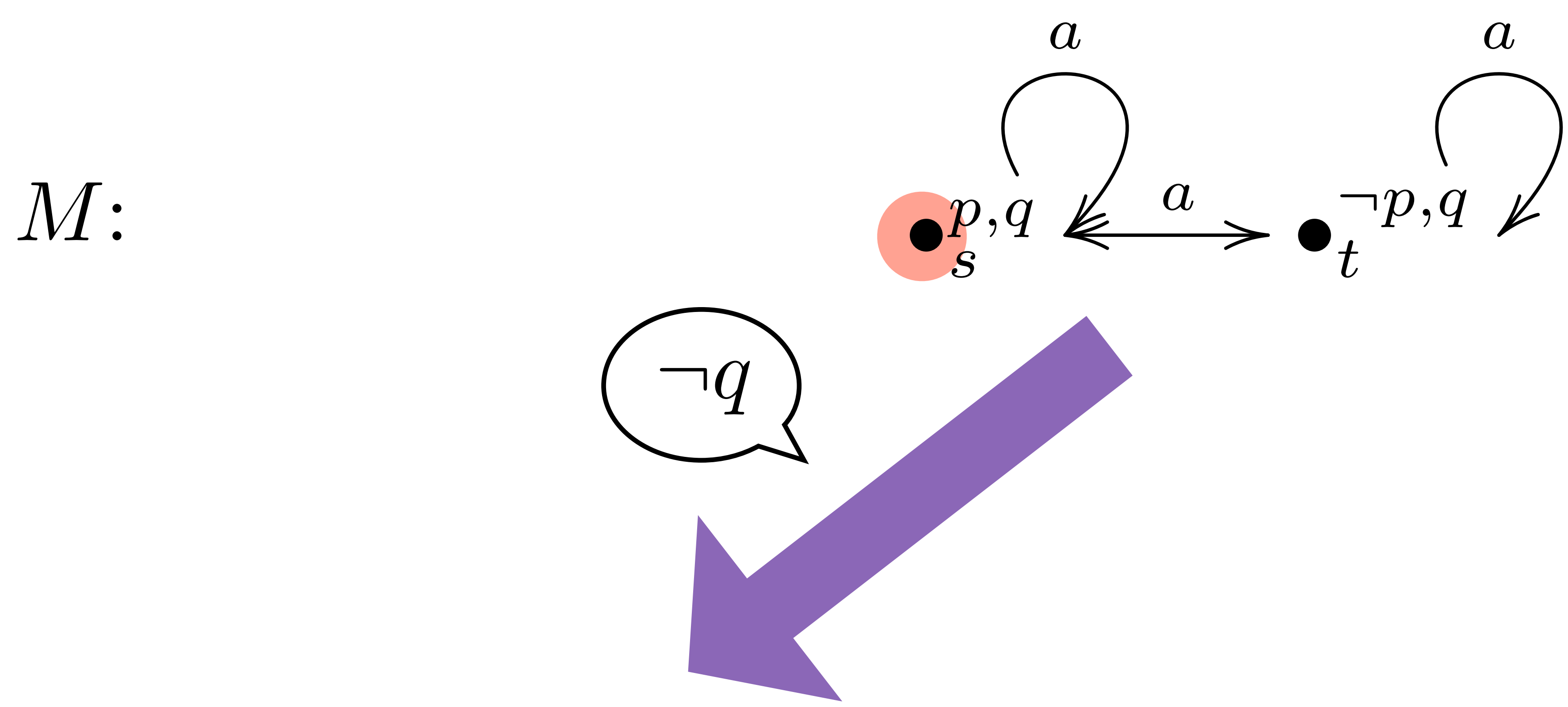


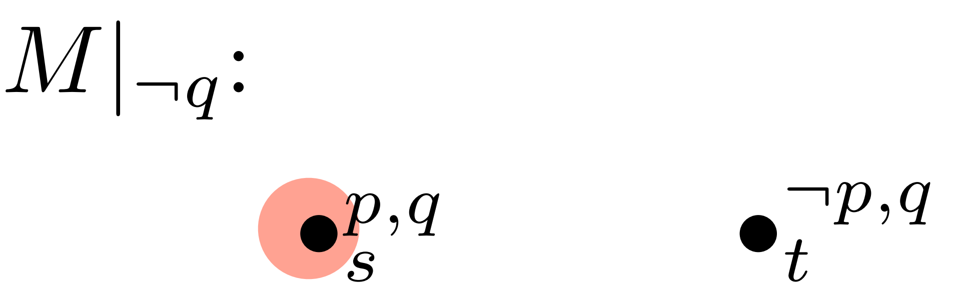


$M|_{\neg q}$:

Believable lie:

Unbelievable lie



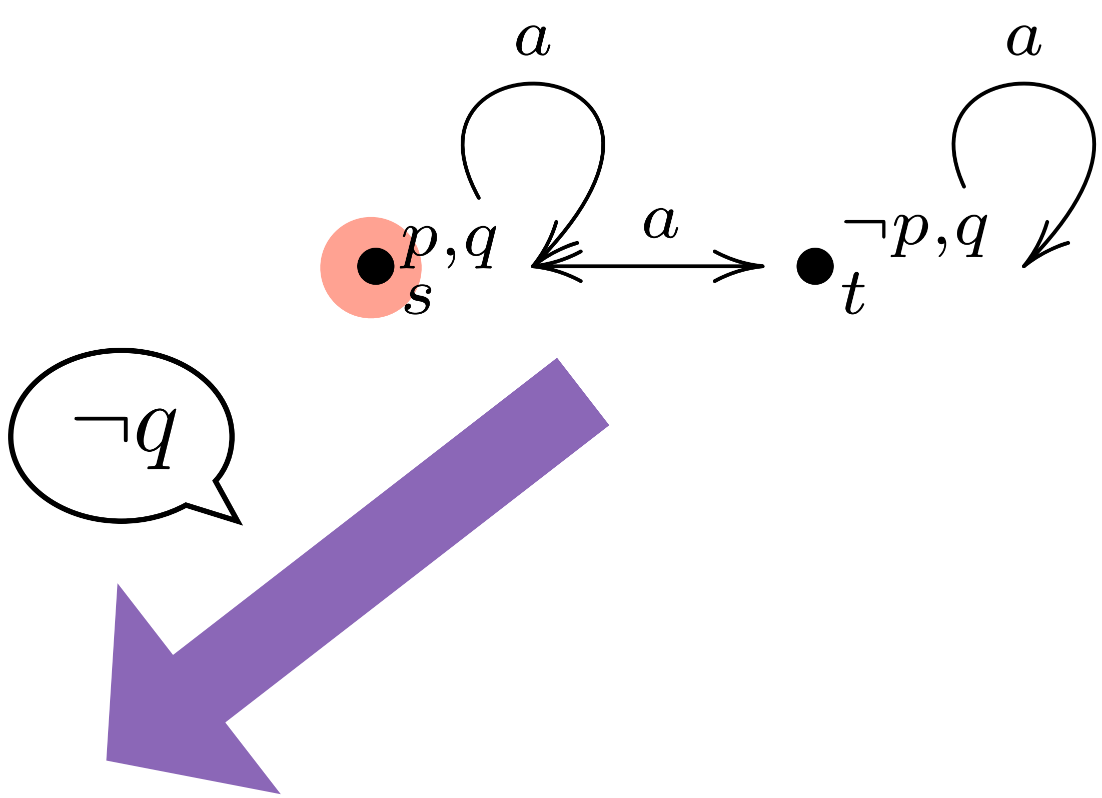


Believable lie:

$M|_{\phi}, s \models \neg \bigvee_{i \in Ag} B_i \bot$

Unbelievable lie

M:



$M|_{\neg q}$:

Believable lie:

$$M|_{\phi}, s \models \neg \bigvee_{i \in Ag} B_i \bot$$

$$\Leftrightarrow M, s \models \bigwedge_{i \in Ag} \hat{B}_i \phi$$

Models of lying

Already seen:

- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Models of lying

Already seen:
- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:

Models of lying

Already seen:

- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:

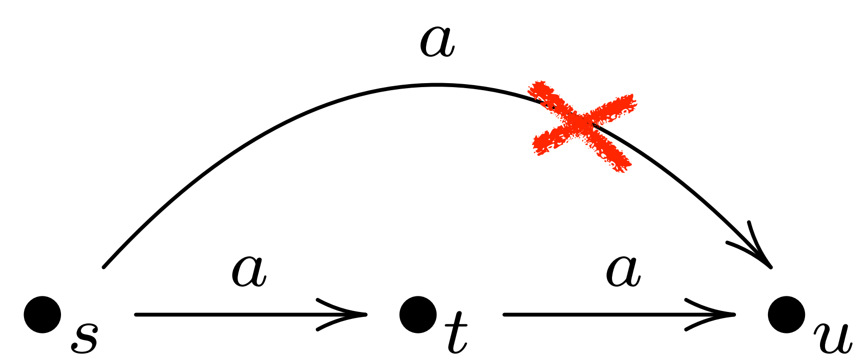
$$\begin{tikzcd} \bullet_s \arrow[r, "a"] \arrow[rr, bend left, "a"] & \bullet_t \arrow[r, "a"] & \bullet_u \end{tikzcd}$$

Models of lying

Already seen:

- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:

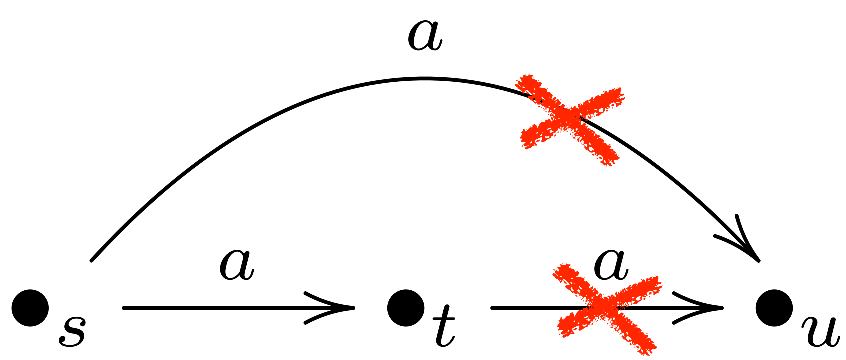


Models of lying

Already seen:

- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:

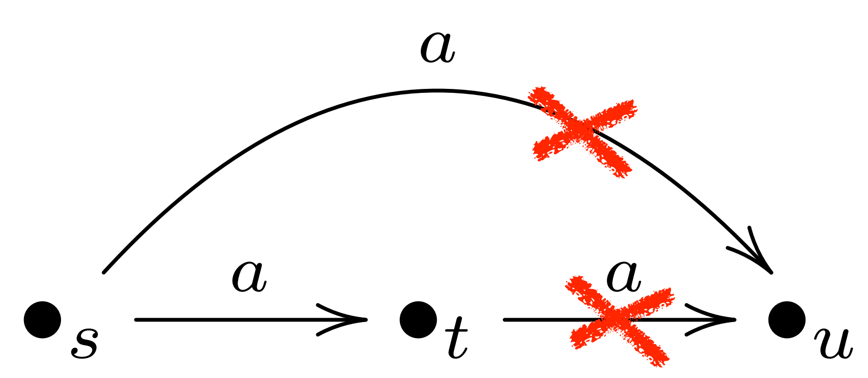


Models of lying

Already seen:

- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:

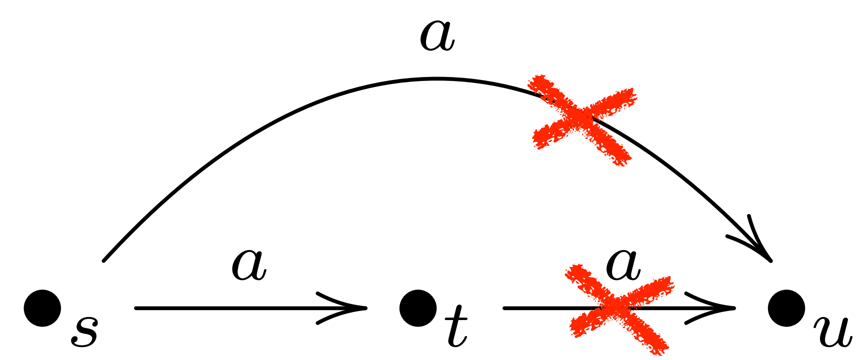


Preservation of Euclidicity:

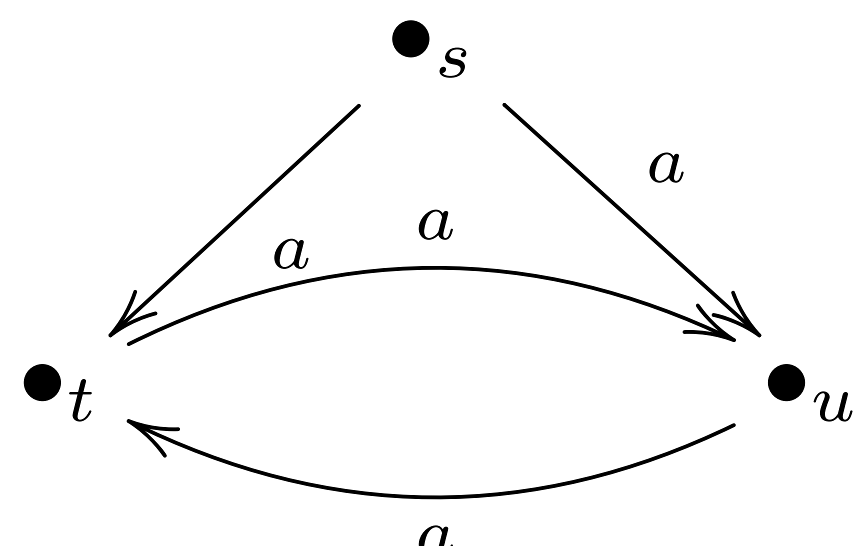
Models of lying

Already seen:
- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:



Preservation of Euclidicity:

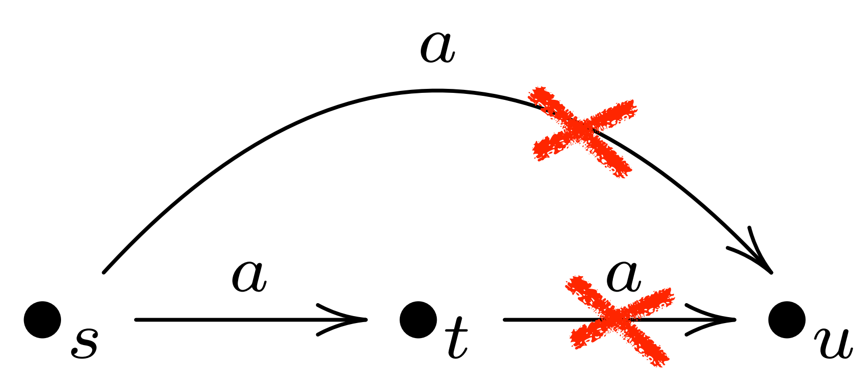


Models of lying

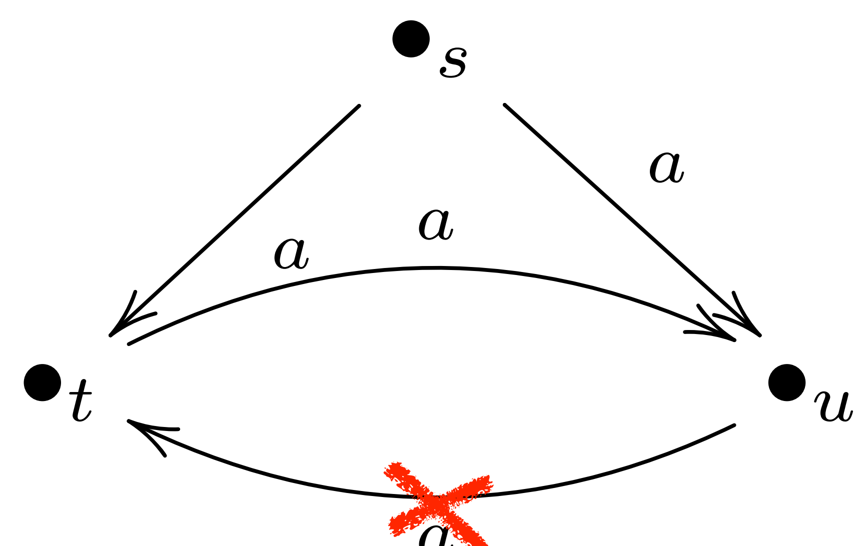
Already seen:

- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:



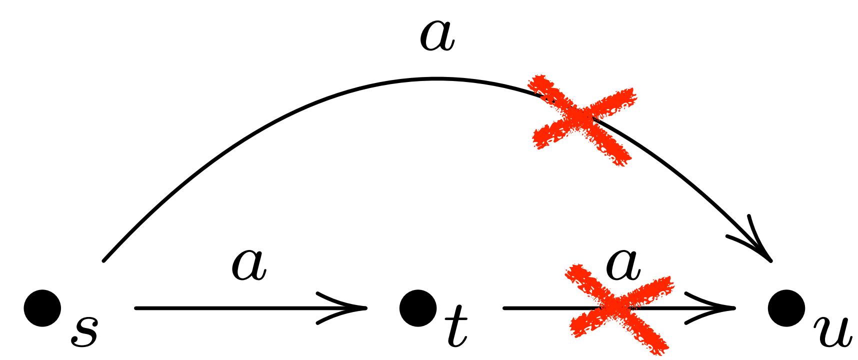
Preservation of Euclidicity:



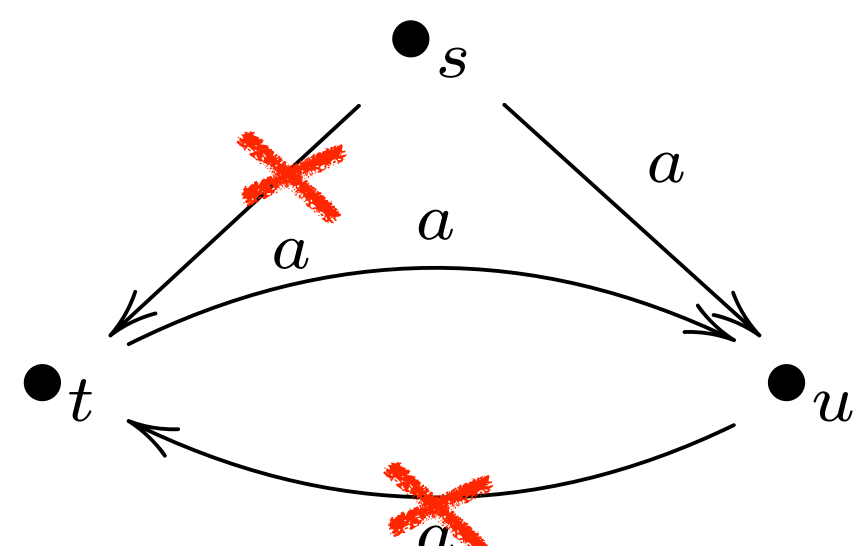
Models of lying

Already seen:
- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:



Preservation of Euclidicity:

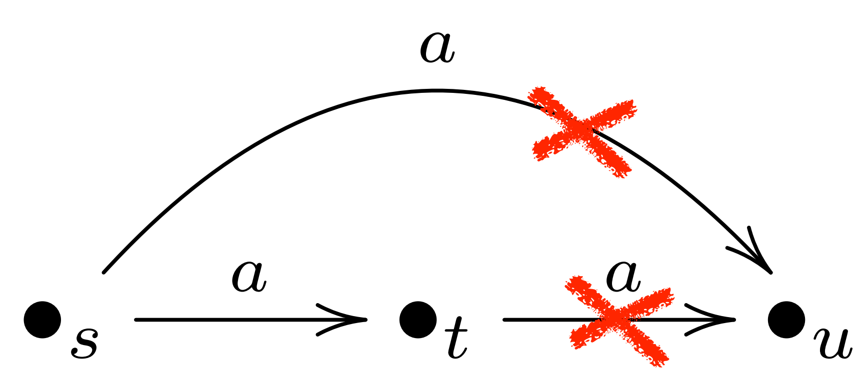


Models of lying

Already seen:

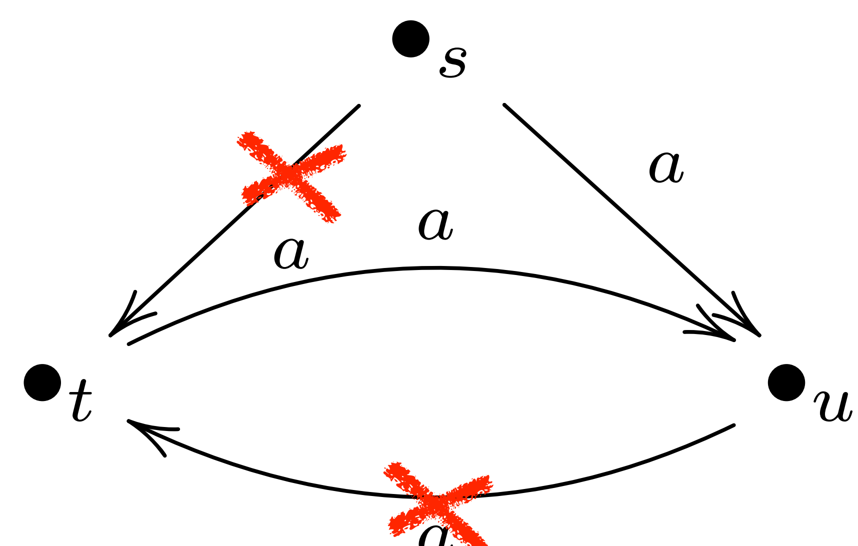
- reflexivity is not preserved under lying
- seriality preserved only for believable lies

Preservation of transitivity:



Models of lying are K45 models, or KD45 models if we only allow believable lies

Prese… city:



True lies: from the outside

$$\phi \text{ is a } \textcolor{red}{\text{true lie in } M, s} \text{ iff } M, s \models \neg\phi \text{ and } M|_{\phi}, s \models \phi$$

True lies: from the outside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \phi$

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

True lies: from the outside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \phi$

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

ϕ is a true lie iff $\forall M \forall s : \; M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$

True lies: from the outside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \phi$

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

ϕ is a true lie iff $\forall M \forall s : (M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot)$

True lies: from the outside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \phi$

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

ϕ is a true lie iff $\forall M \forall s : (M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot)$

In model class K(D)45

Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_{\phi}, s \models \phi$

Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

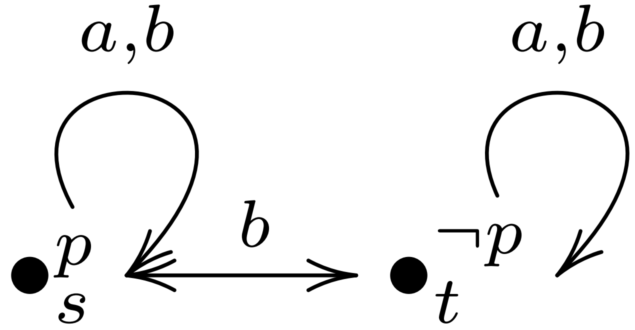
$\phi_0 = p \wedge B_b p$

Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

$\phi_0 = p \wedge B_b p$

M_0:



Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_{\phi}, s \models \phi$

$$\phi_0 = p \wedge B_b p$$

$$M_0\colon \quad \begin{tikzcd} \bullet^{p}_{s} \arrow[loop above, "a{,}b"] \arrow[r, leftrightarrow, "b"] & \bullet^{\neg p}_{t} \arrow[loop above, "a{,}b"] \end{tikzcd}$$

$$M_0|_{\phi_0}\colon \quad \bullet^{p}_{s} \qquad\qquad \bullet^{\neg p}_{t}$$

Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

$\phi_0 = p \wedge B_b p$

$$M_0\colon \quad \begin{tikzcd} \bullet_s^{p} \arrow[loop above, "a{,}b"] \arrow[r, leftrightarrow, "b"] & \bullet_t^{\neg p} \arrow[loop above, "a{,}b"] \end{tikzcd}$$

$$M_0|_{\phi_0}\colon \quad \bullet_s^{p} \qquad\qquad \bullet_t^{\neg p}$$

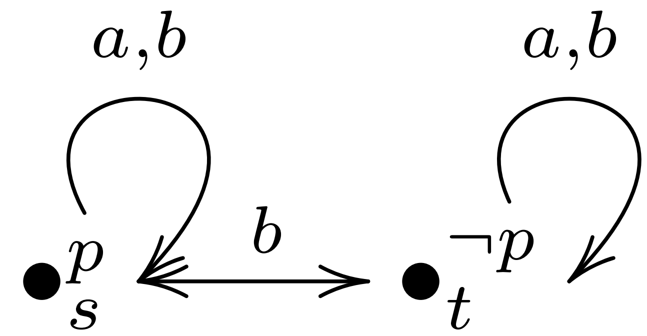
ϕ_0 is a true lie in M_0, s

Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

$\phi_0 = p \wedge B_b p$

M_0:



$M_0|_{\phi_0}$:

ϕ_0 is a true lie in M_0, s

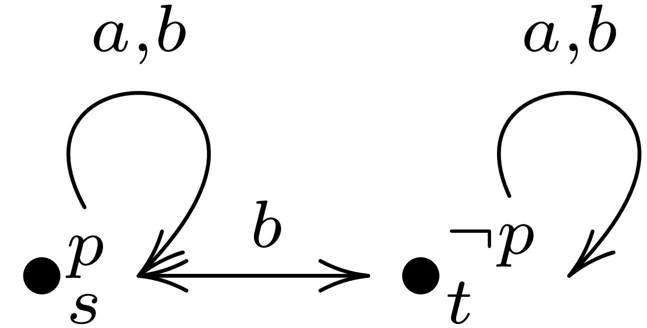
ϕ_0 is not a true lie in M_0, t

Example

$$\phi \text{ is a true lie in } M, s \text{ iff } M, s \models \neg\phi \text{ and } M|_{\phi}, s \models \phi$$

$\phi_0 = p \wedge B_b p$

M_0:



$M_0|_{\phi_0}$:

ϕ_0 is a true lie in M_0, s

ϕ_0 is not a true lie in M_0, t

ϕ_0 is not a believable true lie in M_0, s

Example

ϕ is a true lie iff $\forall M \forall s: \; M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$
believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

$$\phi = \neg B_b p$$

Example

ϕ is a true lie iff $\forall M \forall s: \; M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$
believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

$$\phi = \neg B_b p$$

- ϕ is not a true lie

Example

ϕ is a true lie iff $\forall M \forall s: \; M, s \models \neg\phi \quad \Rightarrow \quad M|_{\phi}, s \models \phi$

believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

$$\phi = \neg B_b p$$

- ϕ is not a true lie
- ϕ is a believable true lie

Example

ϕ is a true lie iff $\forall M \forall s : \; M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$
believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

$$\phi = \neg B_b p$$

- ϕ is not a true lie
- ϕ is a believable true lie
- ϕ is a trivially believable true lie: $\neg\phi \wedge \hat{B}_b\phi$ is inconsistent

Example

ϕ is a true lie iff $\forall M \forall s: \; M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$

believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

Example

ϕ is a true lie iff $\forall M \forall s: \;\; M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$
believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

$$\phi = B_b p$$

Example

ϕ is a true lie iff $\forall M \forall s: \;\; M, s \models \neg\phi \quad \Rightarrow \quad M|_\phi, s \models \phi$
believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

$$\phi = B_b p$$

- ϕ is a true lie

Example

ϕ is a true lie iff $\forall M \forall s : \; M, s \models \neg\phi \quad \Rightarrow \quad M|_{\phi}, s \models \phi$

believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

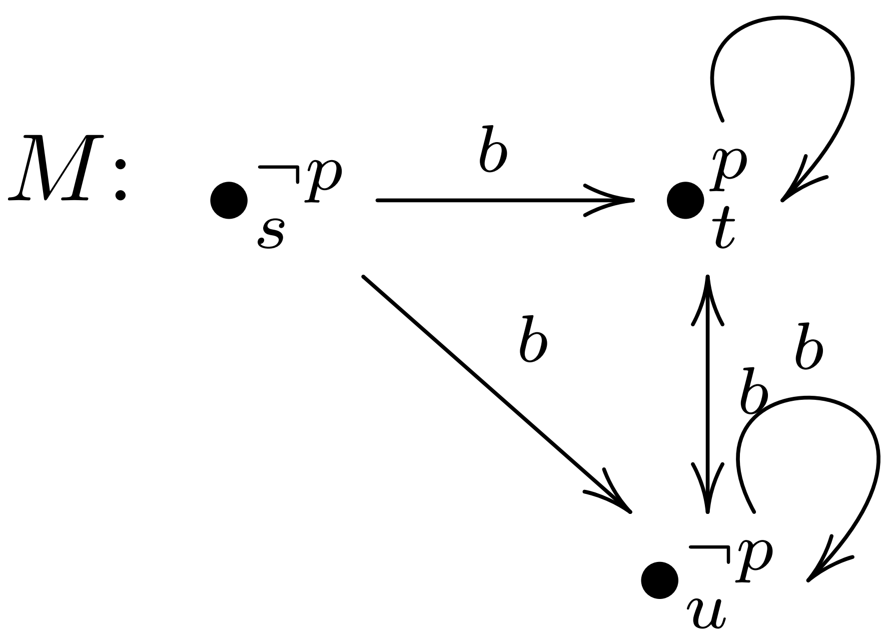
$$\phi = B_b p$$

- ϕ is a true lie
- ϕ is a trivially believable true lie

Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models_b \phi$

$\phi = p \vee B_b p$



Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models_b \phi$

$\phi = p \vee B_b p$

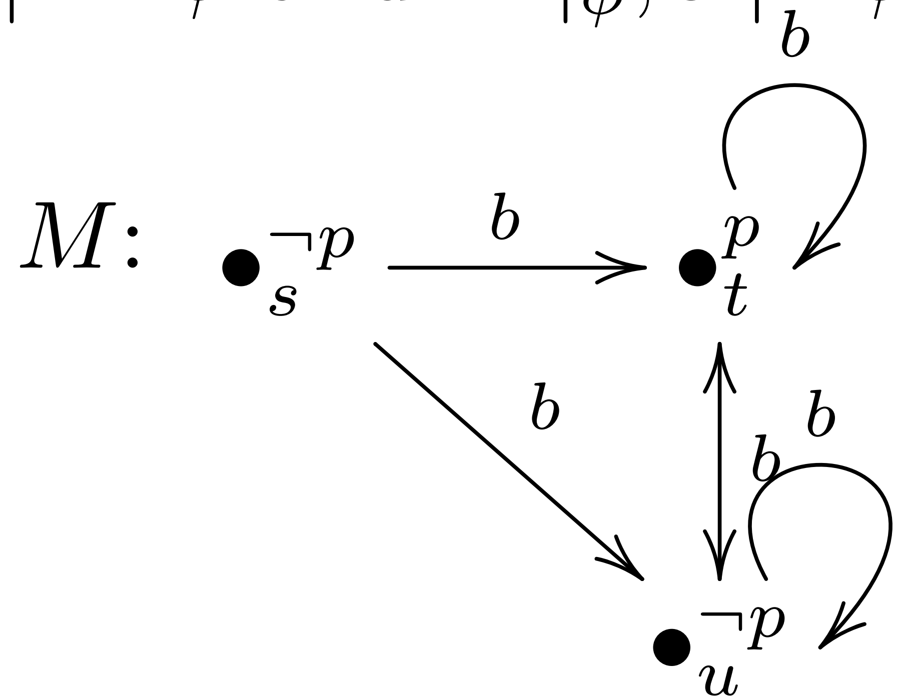
$$\begin{tikzcd}
M\colon & \bullet^{\neg p}_{s} \arrow[r, "b"] \arrow[dr, "b"] & \bullet^{p}_{t} \arrow[loop above, "b"] \arrow[d, shift left, "b"] \\
 & & \bullet^{\neg p}_{u} \arrow[loop right, "b"] \arrow[u, shift left, "b"]
\end{tikzcd}$$

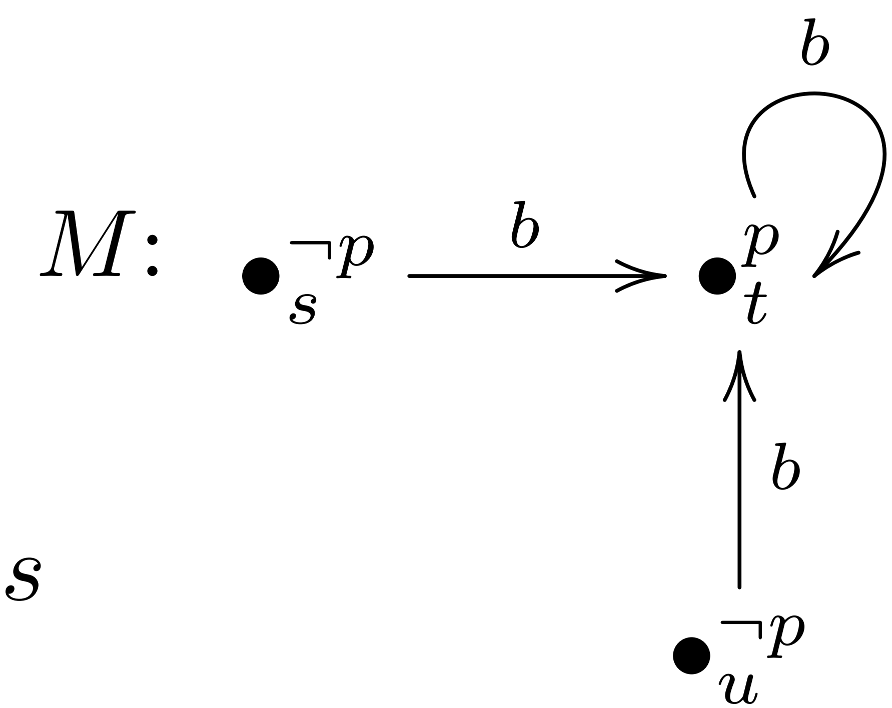
$$\begin{tikzcd}
M\colon & \bullet^{\neg p}_{s} \arrow[r, "b"] & \bullet^{p}_{t} \arrow[loop above, "b"] \\
 & & \bullet^{\neg p}_{u} \arrow[u, "b"]
\end{tikzcd}$$

Example

ϕ is a true lie in M, s iff $M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

$\phi = p \vee B_b p$





ϕ is a believable true lie in M, s

Example

ϕ is a true lie iff $\forall M \forall s : \; M, s \models \neg\phi \quad \Rightarrow \quad M|_{\phi}, s \models \phi$
believable and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot$

$$\phi = p \vee B_b p$$

- ϕ is a true lie
- ϕ is not a trivially believable true lie

Relations to (un)successful updates

True lie in M, s: $\quad M, s \models \neg\phi$ and $M|_{\phi}, s \models \phi$

Relations to (un)successful updates

True lie in M, s: $\quad M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

Successful update in M, s: $\quad M, s \models \phi$ and $M|_\phi, s \models \phi$

Relations to (un)successful updates

True lie in M, s: $\quad M, s \models \neg\phi$ and $M|_\phi, s \models \phi$

Successful update in M, s: $\quad M, s \models \phi$ and $M|_\phi, s \models \phi$

Unsuccessful update in M, s: $\quad M, s \models \phi$ and $M|_\phi, s \models \neg\phi$

Other Moorean definitions

$$
\begin{array}{lcccc}
\text{Self-refuting truth:} & \forall M, s & M, s \models \phi & \Rightarrow & M|_\phi, s \models \neg\phi \\
\text{True lie:} & \forall M, s & M, s \models \neg\phi & \Rightarrow & M|_\phi, s \models \phi \\
\text{Successful formula:} & \forall M, s & M, s \models \phi & \Rightarrow & M|_\phi, s \models \phi \\
\text{Impossible lie:} & \forall M, s & M, s \models \neg\phi & \Rightarrow & M|_\phi, s \models \neg\phi
\end{array}
$$

Other Moorean definitions

$$\begin{array}{llcll} \text{Self-refuting truth:} & \forall M, s & M, s \models \phi & \Rightarrow & M|_{\phi}, s \models \neg\phi \\ \text{True lie:} & \forall M, s & M, s \models \neg\phi & \Rightarrow & M|_{\phi}, s \models \phi \\ \text{Successful formula:} & \forall M, s & M, s \models \phi & \Rightarrow & M|_{\phi}, s \models \phi \\ \text{Impossible lie:} & \forall M, s & M, s \models \neg\phi & \Rightarrow & M|_{\phi}, s \models \neg\phi \end{array}$$

$$\phi = p \wedge \neg B_b p$$

- Unsuccessful
- Self-refuting

Syntactic characterisation of true lies

- Exactly which formulae are (believable) true lies?
- We give a syntactic characterisation of believable true lies for the single-agent case
- The technique is based on Holliday and Icard (AiML 2010), who characterise the unsuccessful and self-refuting formulas (also in the single-agent case)

Characterisation: preliminaries

Every KD45 formula is equivalent to one on normal form: a disjunction of conjunctions of the form

$$\delta = \alpha \wedge \Box\beta_1 \wedge \ldots \wedge \Box\beta_n \wedge \Diamond\gamma_1 \wedge \ldots \wedge \Diamond\gamma_m$$

where α and γ_i are conjunctions of literals and β_i is a disjunction of literals.

Characterisation: preliminaries

Every KD45 formula is equivalent to one on normal form: a disjunction of conjunctions of the form

$$\delta = \alpha \wedge \Box\beta_1 \wedge \ldots \wedge \Box\beta_n \wedge \Diamond\gamma_1 \wedge \ldots \wedge \Diamond\gamma_m$$

where α and γ_i are conjunctions of literals and β_i is a disjunction of literals.

Clarity (Holliday and Icard) Given a conjunction or disjunction χ of literals, $L(\chi)$ denotes the set of literals. $L(\chi)$ is open iff no literal in $L(\chi)$ is the negation of any other. A conjunction $\delta = \alpha \wedge \Box\beta_1 \wedge \ldots \wedge \Box\beta_n \wedge \Diamond\gamma_1 \wedge \ldots \wedge \Diamond\gamma_m$ on normal form is clear iff (i) $L(\alpha)$ is open; (ii) there is an open set of literals $\{l_1, \ldots, l_n\}$ with $l_i \in L(\beta_i)$; and (iii) for every γ_k there is a set of literals $\{l_1, \ldots, l_n\}$ with $l_i \in L(\beta_i)$ such that $\{l_1, \ldots, l_n\} \cup L(\gamma_k)$ is open. A disjunction on normal form is clear iff at least one of the disjuncts are clear.

Characterisation: main result (single agent)

Definition 1. *A formula ϕ on normal form is an unsuccessful lie iff there exists sets S and T of disjuncts of ϕ such that every $\theta \in T$ has a conjunct $\Box\beta_\theta$ such that any normal form of*

$$\chi = \neg\phi \land \Diamond\phi \land \chi_1 \land \chi_2 \land \chi_3$$

is clear, where

$$\chi_1 = \bigwedge_{\theta \in T} t(\theta) \land \bigwedge_{\theta \notin T} \neg t(\theta) \qquad t(\theta) = \theta^\alpha \land \bigwedge_{\Diamond\gamma \text{ in } \theta} \bigvee_{\sigma \in S} \Diamond(\sigma^\alpha \land \gamma)$$

$$\chi_2 = \bigwedge_{\sigma \in S} \sigma^{\Box\Diamond} \land \bigwedge_{\sigma \notin S} \neg\sigma^{\Box\Diamond} \qquad \chi_3 = \bigwedge_{\theta \in T} \bigvee_{\sigma \in S} \Diamond(\sigma^\alpha \land \sim \beta_\theta)$$

ϕ is an unsuccessful lie iff any normal form of ϕ is an unsuccessful lie.

$$\delta = \alpha \land \Box\beta_1 \land \ldots \land \Box\beta_n \land \Diamond\gamma_1 \land \ldots \land \Diamond\gamma_m$$

Characterisation: main result (single agent)

Definition 1. *A formula ϕ on normal form is an unsuccessful lie iff there exists sets S and T of disjuncts of ϕ such that every $\theta \in T$ has a conjunct $\Box\beta_\theta$ such that any normal form of*

$$\chi = \neg\phi \wedge \Diamond\phi \wedge \chi_1 \wedge \chi_2 \wedge \chi_3$$

is clear, where

$$\chi_1 = \bigwedge_{\theta \in T} t(\theta) \wedge \bigwedge_{\theta \notin T} \neg t(\theta) \qquad t(\theta) = \theta^\alpha \wedge \bigwedge_{\Diamond\gamma \text{ in } \theta} \bigvee_{\sigma \in S} \Diamond(\sigma^\alpha \wedge \gamma)$$

$$\chi_2 = \bigwedge_{\sigma \in S} \sigma^{\Box\Diamond} \wedge \bigwedge_{\sigma \notin S} \neg\sigma^{\Box\Diamond} \qquad \chi_3 = \bigwedge_{\theta \in T} \bigvee_{\sigma \in S} \Diamond(\sigma^\alpha \wedge \sim \beta_\theta)$$

ϕ is an unsuccessful lie iff any normal form of ϕ is an unsuccessful lie.

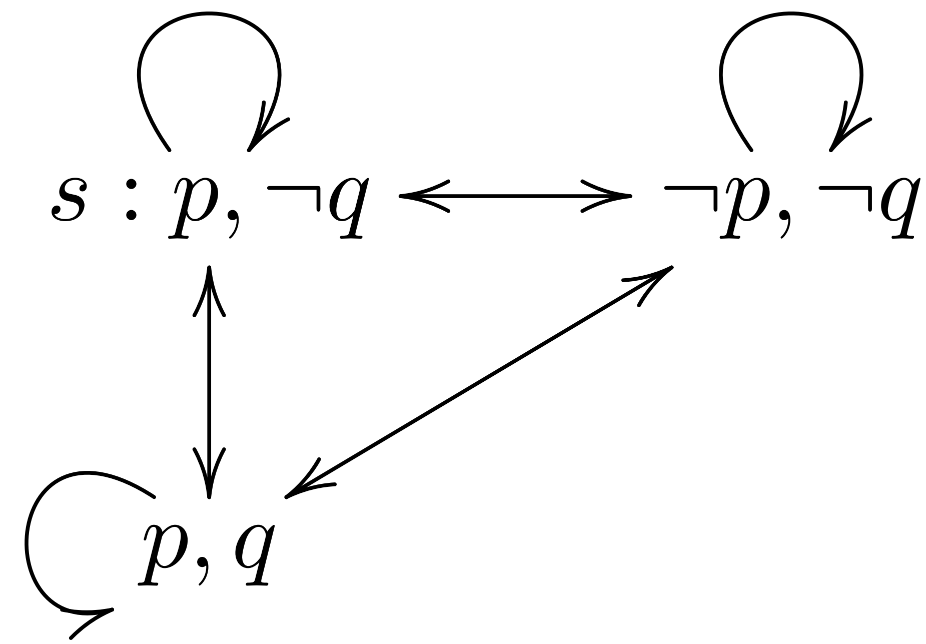
Theorem 1. *A formula ϕ is not a believable true lie if and only if it is an unsuccessful lie.*

$$\delta = \alpha \wedge \Box\beta_1 \wedge \ldots \wedge \Box\beta_n \wedge \Diamond\gamma_1 \wedge \ldots \wedge \Diamond\gamma_m$$

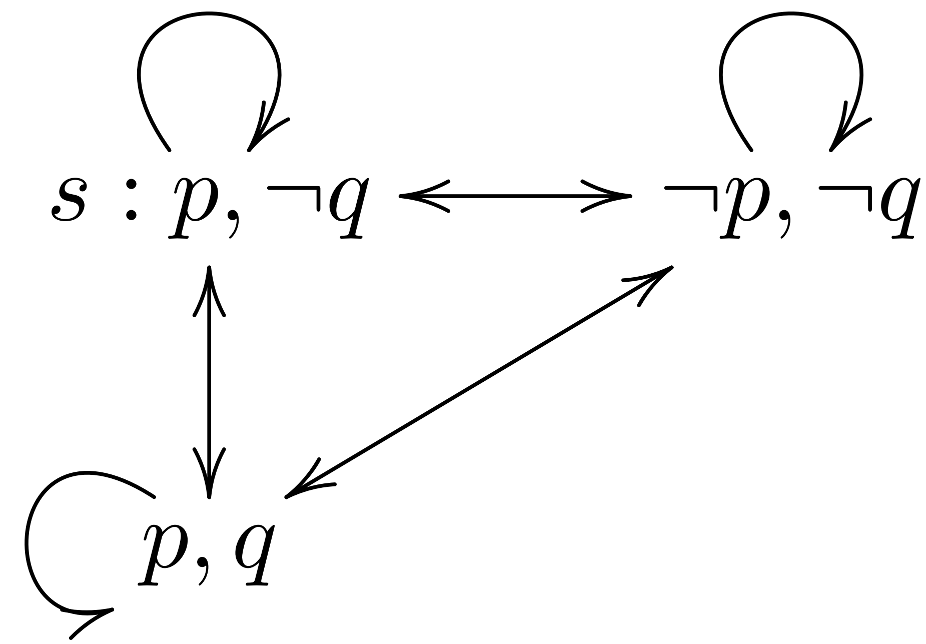
Alternation

Alternation example: true-false-true

$$\phi = (q \vee Bq) \vee (p \wedge \neg Bp)$$



Alternation example: true-false-true



$$\phi = (q \vee Bq) \vee (p \wedge \neg Bp)$$

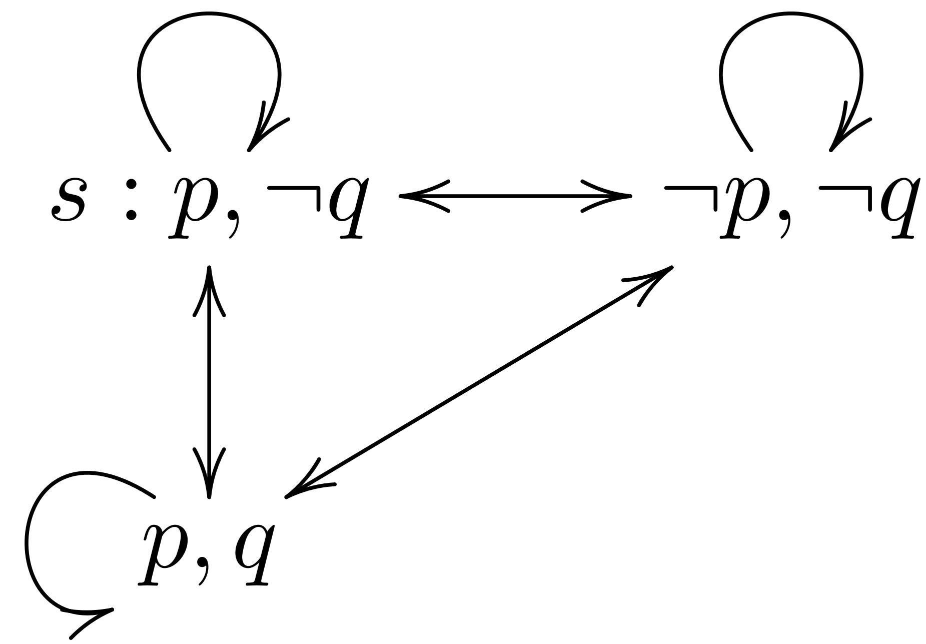
$$M, s \models \phi$$

$$M|_\phi, s \models \neg\phi$$

$$(M|_\phi)|_\phi, s \models \phi$$

Alternation example: true-false-true

$$\phi = (q \vee Bq) \vee (p \wedge \neg Bp)$$



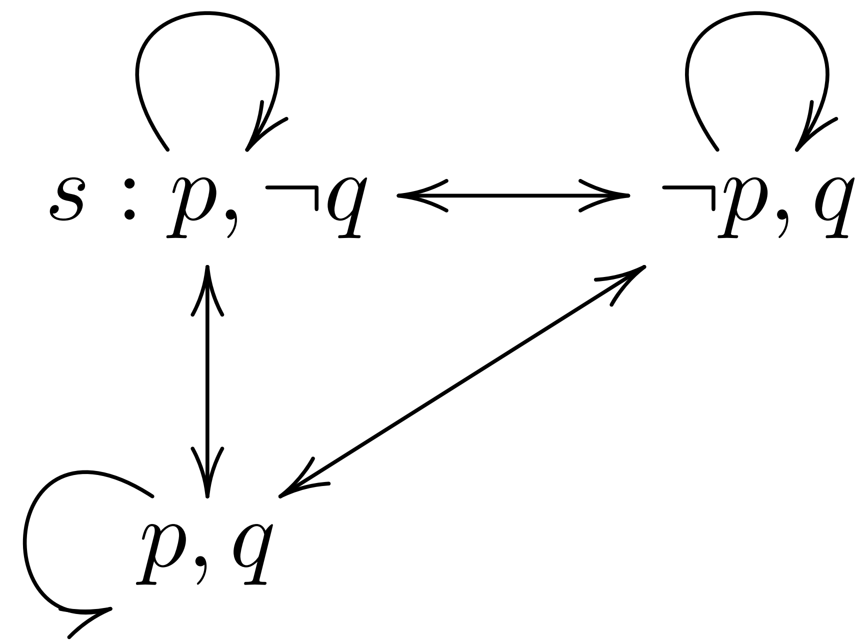
$$M, s \models \phi \qquad M, s \models \hat{B}\phi$$

$$M|_{\phi}, s \models \neg\phi \qquad M|_{\phi}, s \models \hat{B}\phi$$

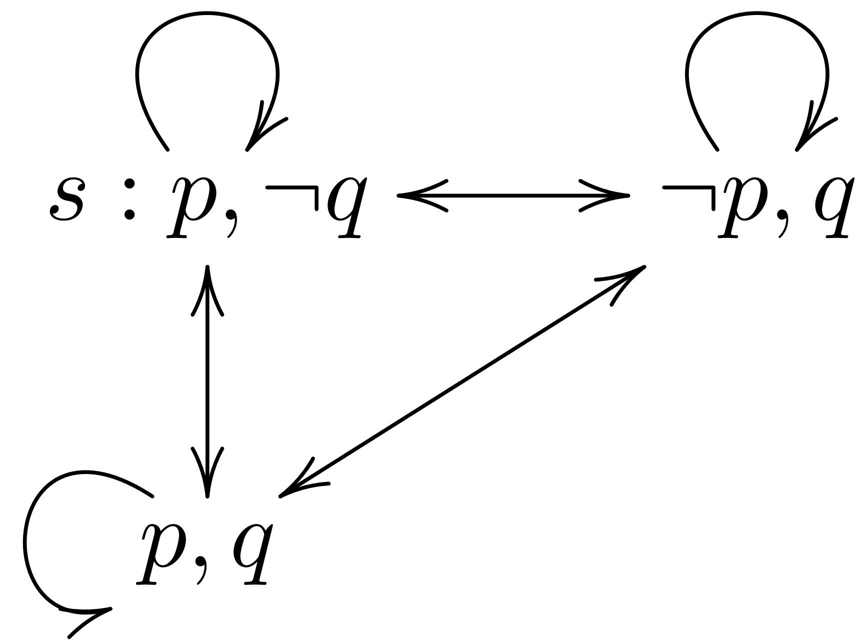
$$(M|_{\phi})|_{\phi}, s \models \phi$$

Alternation example: false-true-false

$$\phi = (q \vee Bq) \wedge ((p \vee \neg Bq) \wedge \neg Bp)$$



Alternation example: false-true-false



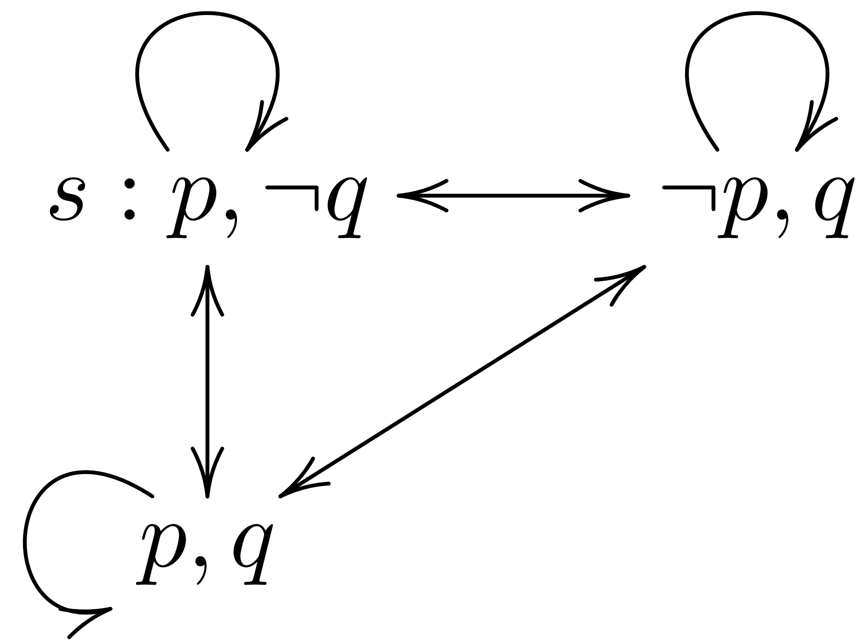
$$\phi = (q \vee Bq) \wedge ((p \vee \neg Bq) \wedge \neg Bp)$$

$$M, s \models \neg\phi$$

$$M|_\phi, s \models \phi$$

$$(M|_\phi)|_\phi, s \models \neg\phi$$

Alternation example: false-true-false



$$\phi = (q \vee Bq) \wedge ((p \vee \neg Bq) \wedge \neg Bp)$$

$$M, s \models \neg\phi \qquad M, s \models \hat{B}\phi$$

$$M|_\phi, s \models \phi \qquad M|_\phi, s \models \hat{B}\phi$$

$$(M|_\phi)|_\phi, s \models \neg\phi$$

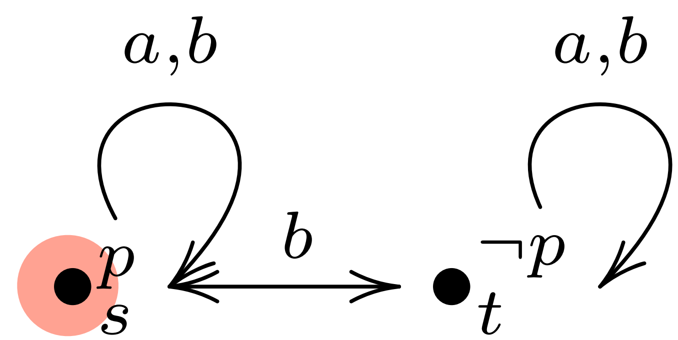
Alternation: open questions

- Do examples exist for every finite alternation sequence?
- If not, how to characterise realisable sequences?
- A stronger version: for which sequences is there a formula that can realise it on *any* model?

True lies from the inside

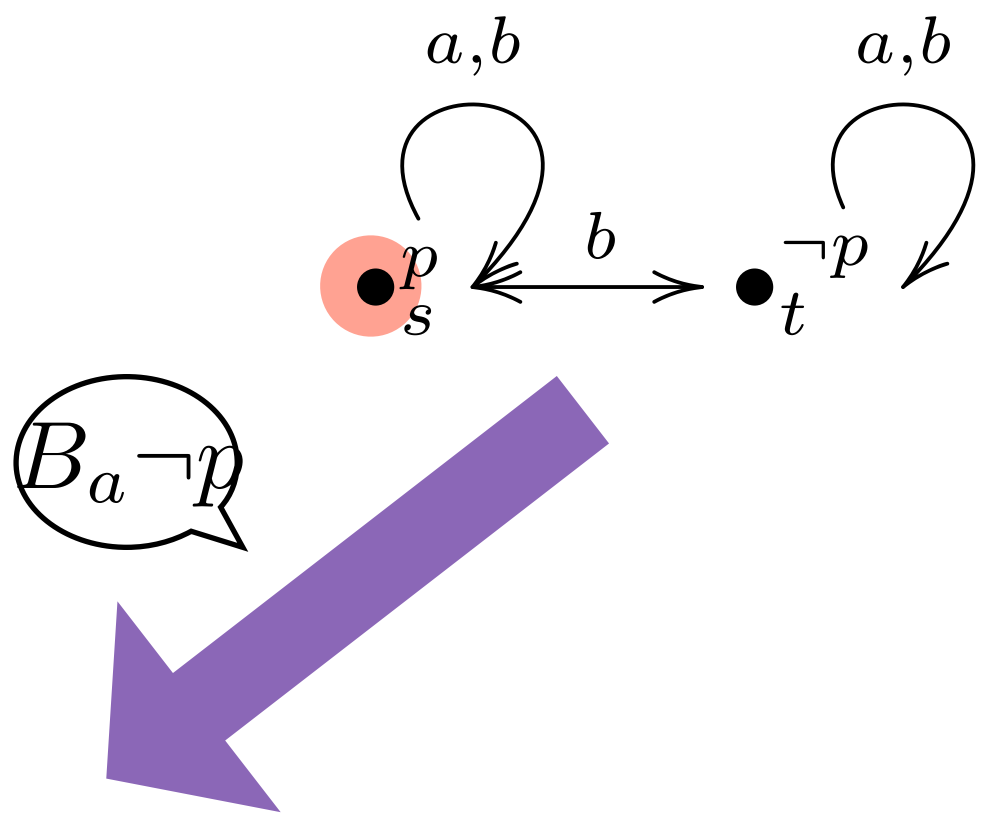
Untruthful announcements by an agent *a* inside the system

M:

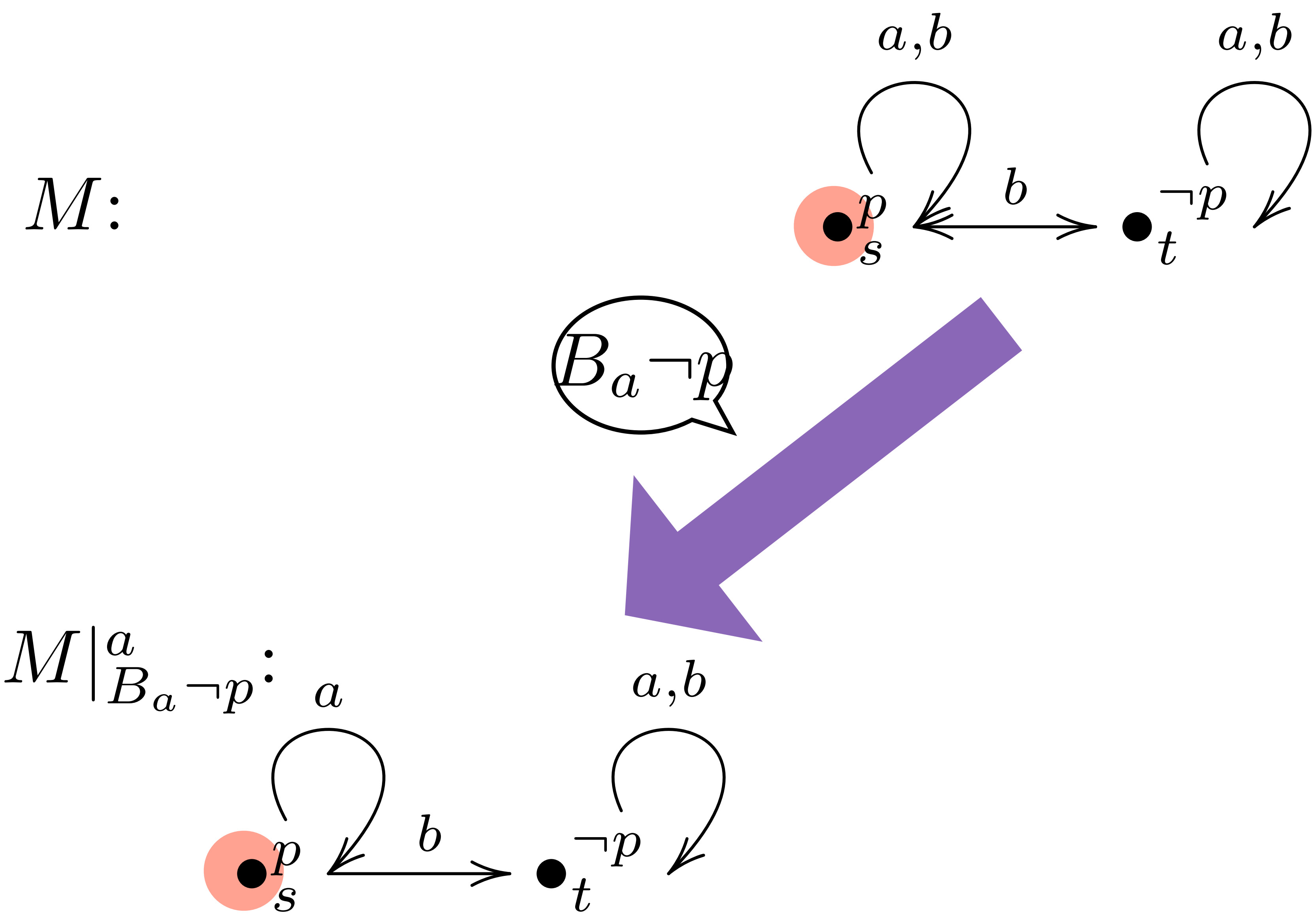


Untruthful announcements by an agent *a* inside the system

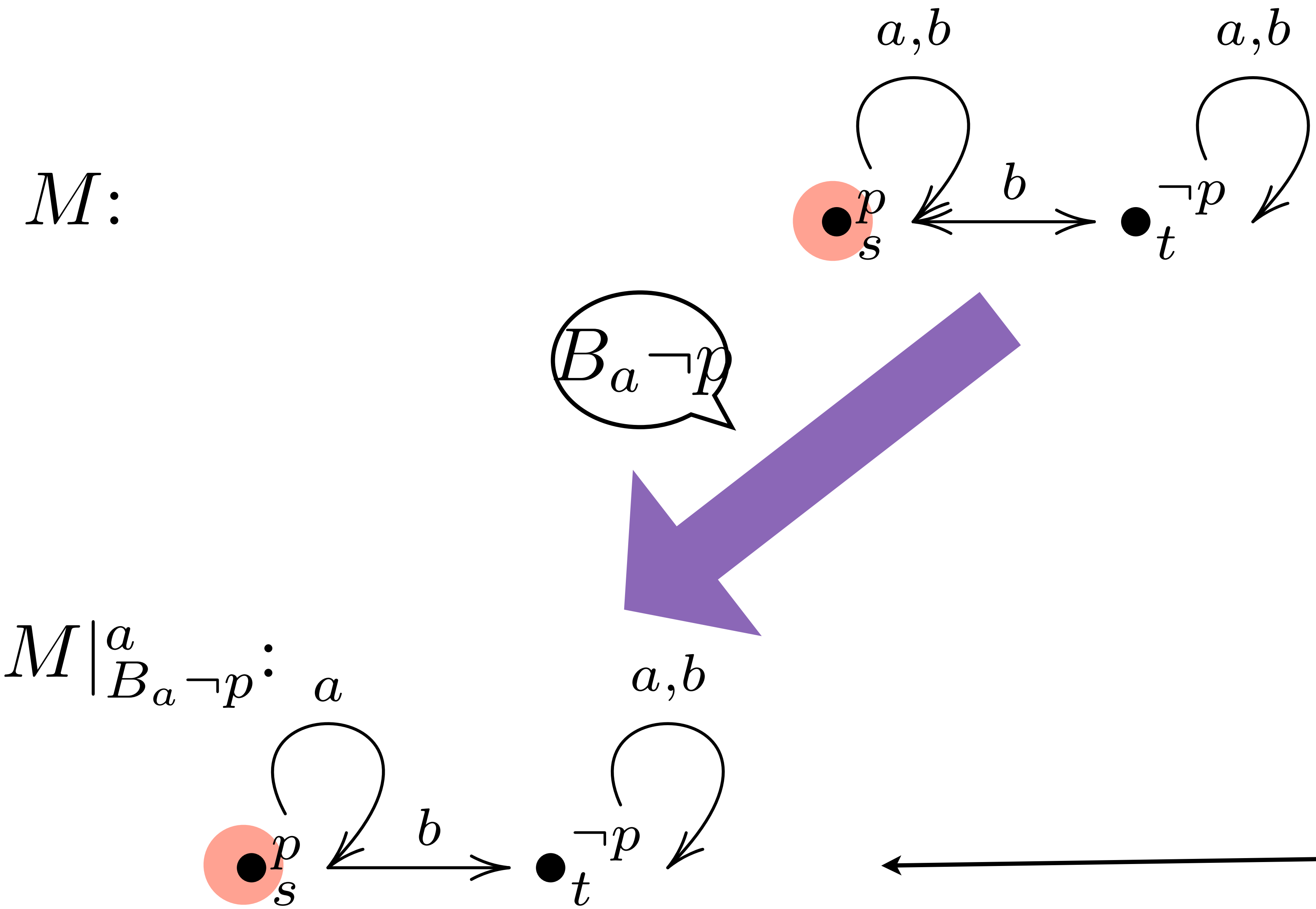
$M:$



Untruthful announcements by an agent *a* inside the system

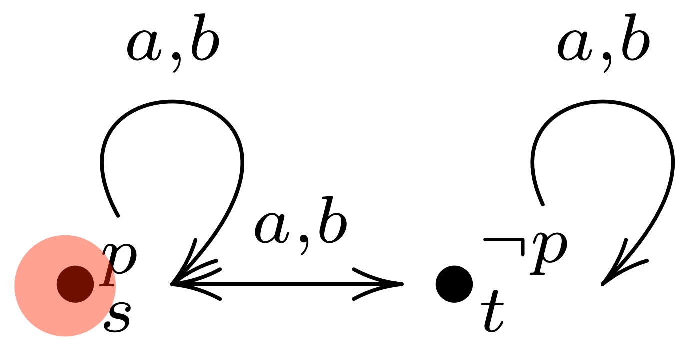


Untruthful announcements by an agent *a* inside the system

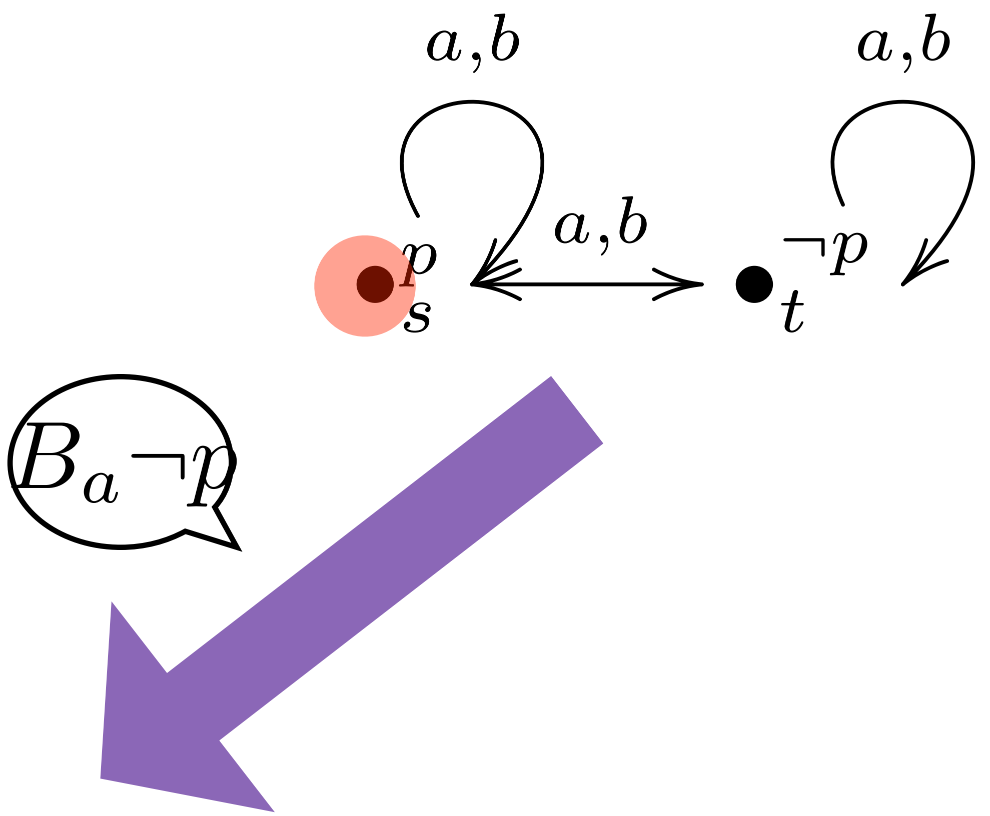


Update obtained by removing links going into states where the announcement is false, for all other agents than *a*

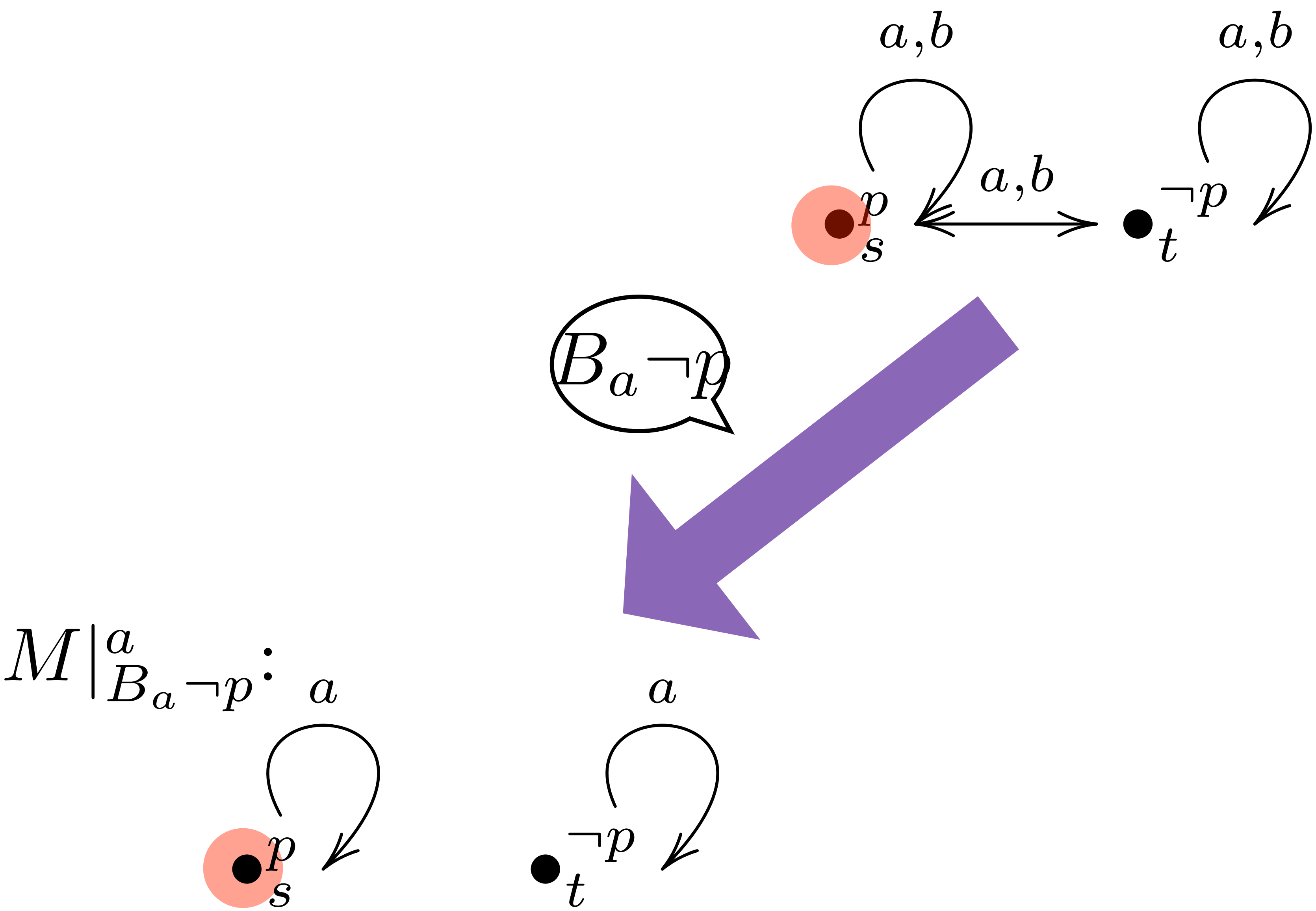
Unbelievable lie



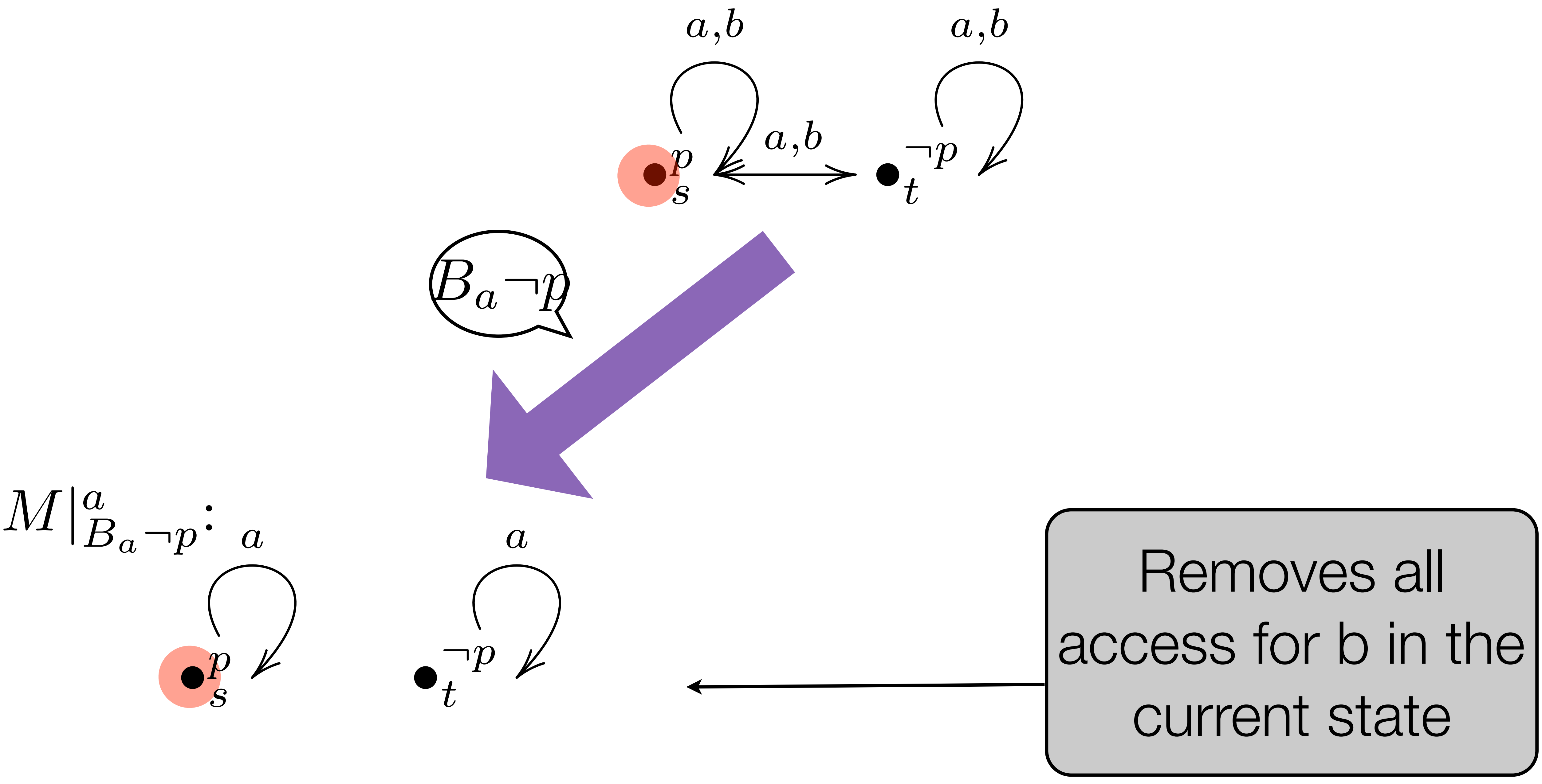
Unbelievable lie



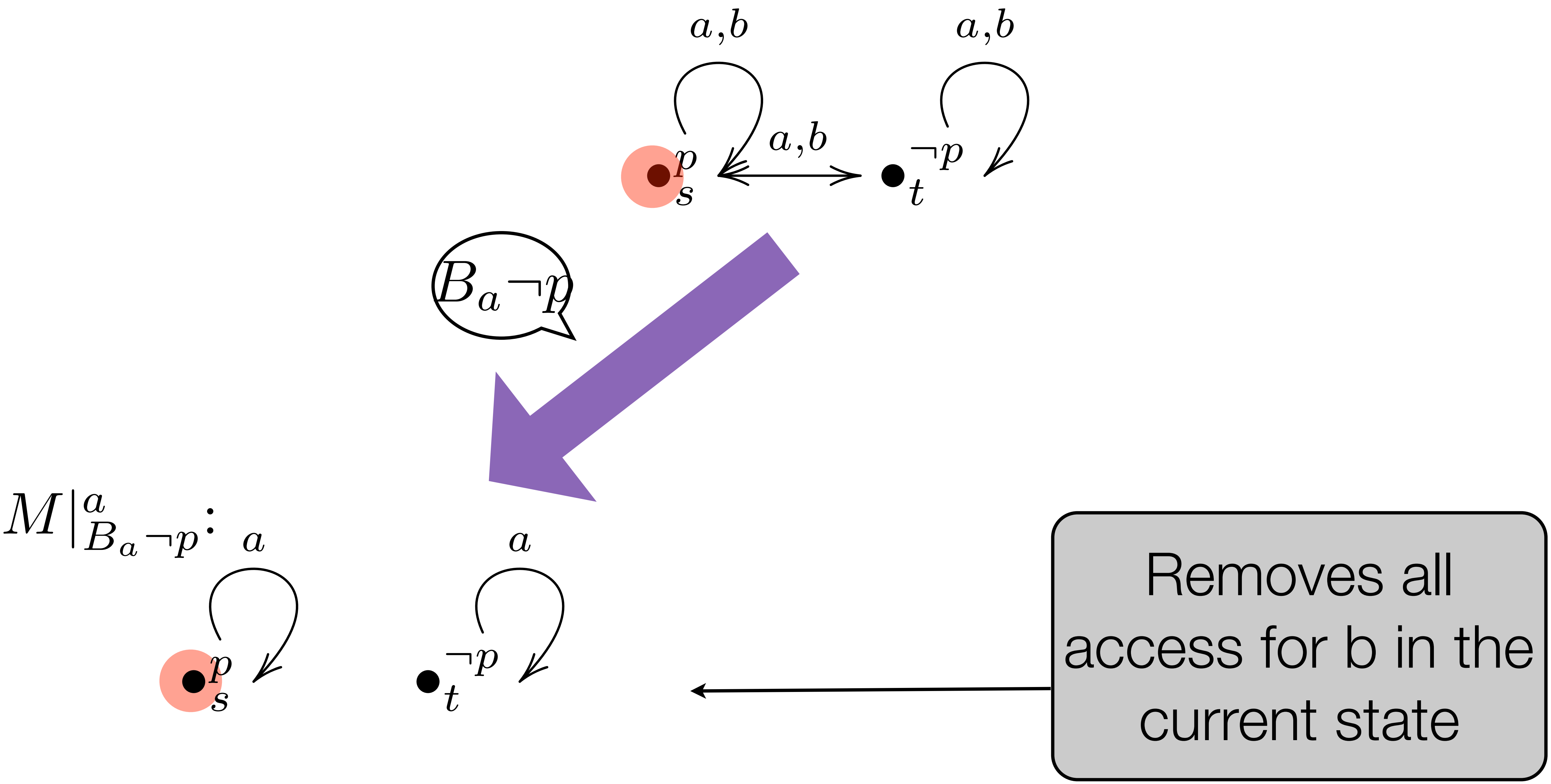
Unbelievable lie



Unbelievable lie

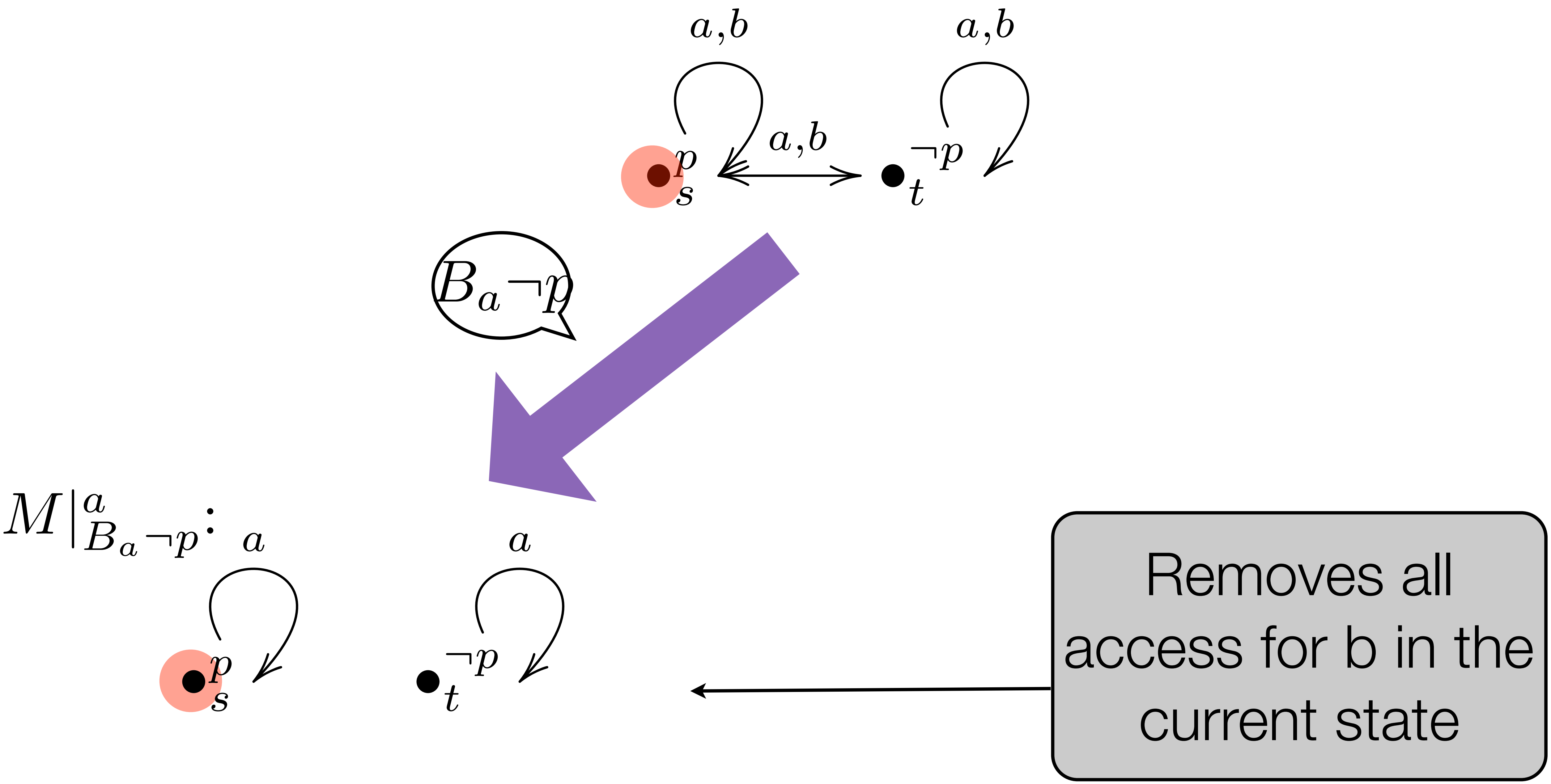


Unbelievable lie



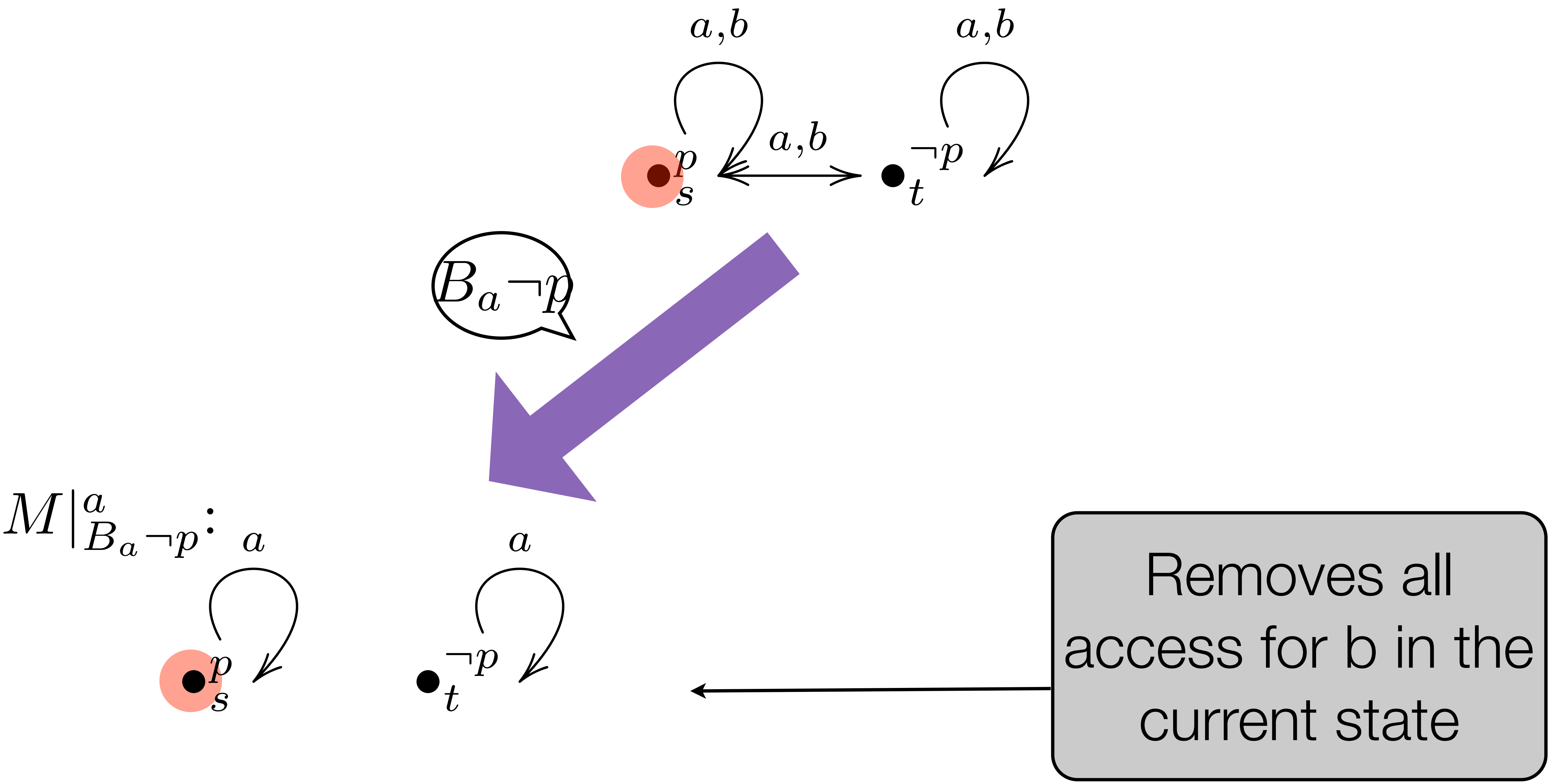
Believable lie:

Unbelievable lie



Believable lie: $\quad M|^a_{B_a \phi}, s \models \neg \bigvee_{i \in Ag} B_i \bot$

Unbelievable lie



Believable lie: $\quad M|^a_{B_a\phi}, s \models \neg \bigvee_{i \in Ag} B_i \bot$

$\Leftrightarrow M, s \models \bigwedge_{i \in Ag} \hat{B}_i B_a \phi$

Lie by agent *a*, possible pre-conditions

$\neg\phi$

Lie by agent *a*, possible pre-conditions

$\neg\phi$

$\neg B_a \phi$

Lie by agent *a*, possible pre-conditions

$\neg\phi$

$\neg B_a\phi$

$B_a\neg\phi$

Lie by agent *a*, possible pre-conditions

$$\neg\phi$$

$$\neg B_a\phi$$

$$B_a\neg\phi$$

$$\neg(B_a\phi \vee B_a\neg\phi)$$

Lie by agent *a*, possible pre-conditions

$\neg\phi$

$\neg B_a \phi$

$B_a \neg\phi$

$\neg(B_a \phi \vee B_a \neg\phi)$

True lie by agent *a*, possible post-conditions

ϕ

True lie by agent *a*, possible post-conditions

ϕ

$B_a\phi$

True lie by agent *a*, possible post-conditions

$\boxed{\phi}$

$B_a\phi$

True lies: from the inside

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^{a}_{B_a \phi}, s \models \phi$

True lies: from the inside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi$

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

True lies: from the inside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi$

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg \phi$ and $M|^a_{B_a \phi}, s \models \phi$

ϕ is a true lie by a iff $\forall M \forall s : \; M, s \models B_a \neg \phi \Rightarrow \quad M|^a_{B_a \phi}, s \models \phi$

True lies: from the inside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi$

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

ϕ is a true lie by a iff $\forall M \forall s : (M, s \models B_a \neg\phi \Rightarrow \quad M|^a_{B_a\phi}, s \models \phi$

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi)$

True lies: from the inside

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi$

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

ϕ is a true lie by a iff $\forall M \forall s : (M, s \models B_a \neg\phi \Rightarrow \quad M|^a_{B_a\phi}, s \models \phi$

believable

and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi)$

In model class K(D)45

Example: from the inside

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

Example: from the inside

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

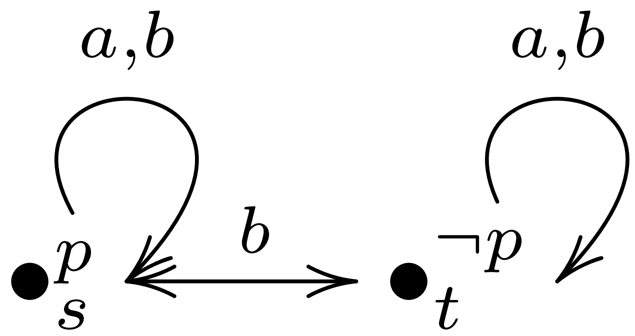
$$\phi_0 = p \wedge B_b p$$

Example: from the inside

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$$\phi_0 = p \wedge B_b p$$

M_0:



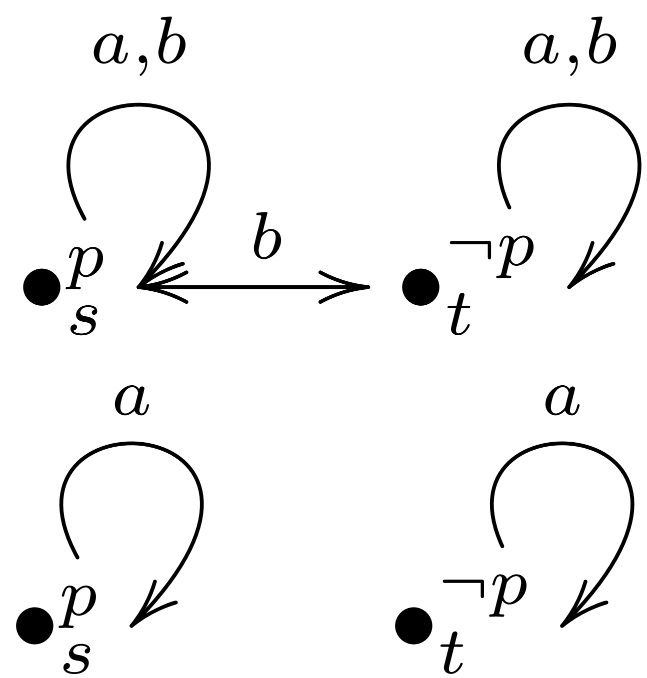
Example: from the inside

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$$\phi_0 = p \wedge B_b p$$

M_0:

$M_0|^a_{B_a\phi_0}$:

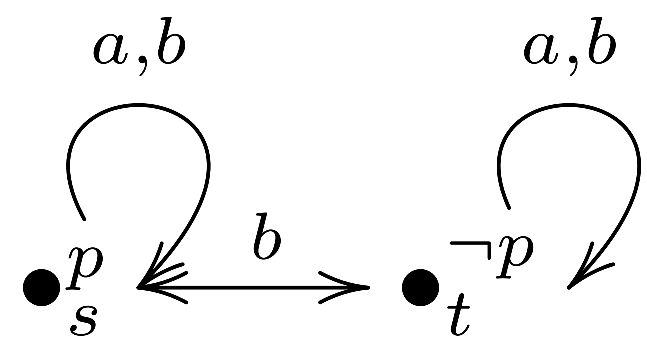


Example: from the inside

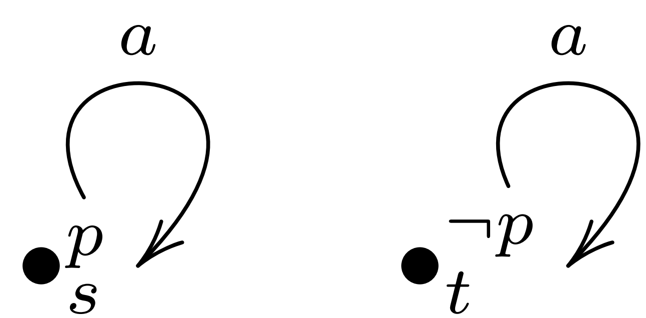
ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$$\phi_0 = p \wedge B_b p$$

M_0:



$M_0|^a_{B_a\phi_0}$:



ϕ_0 is a true lie by a in M_0, s

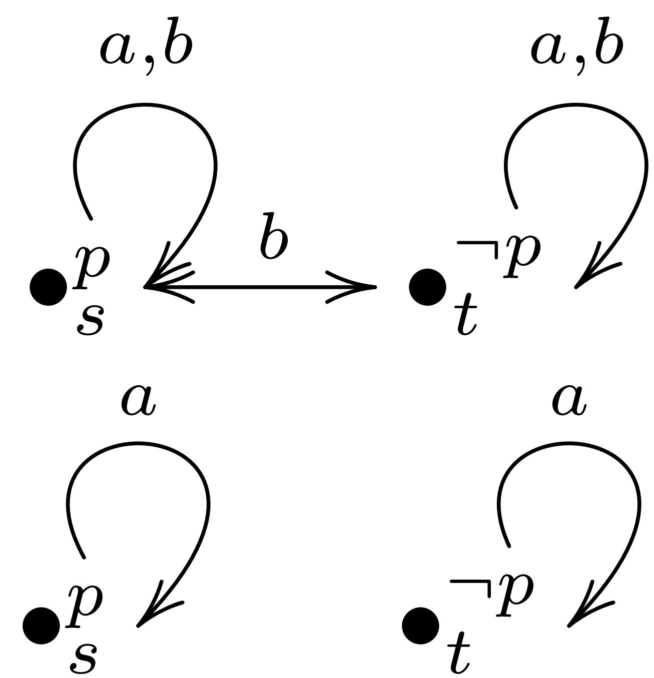
Example: from the inside

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$$\phi_0 = p \wedge B_b p$$

M_0:

$M_0|^a_{B_a\phi_0}$:



ϕ_0 is a true lie by a in M_0, s

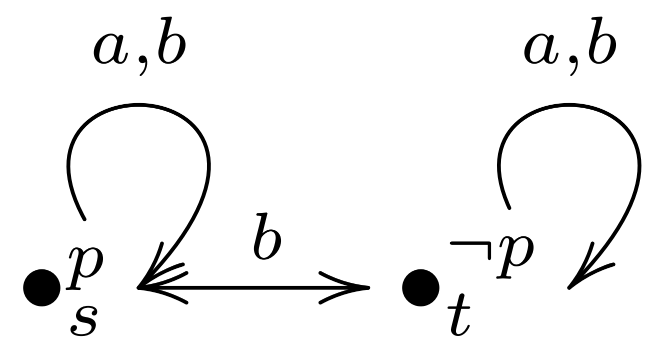
ϕ_0 is not a true lie by a in M_0, t

Example: from the inside

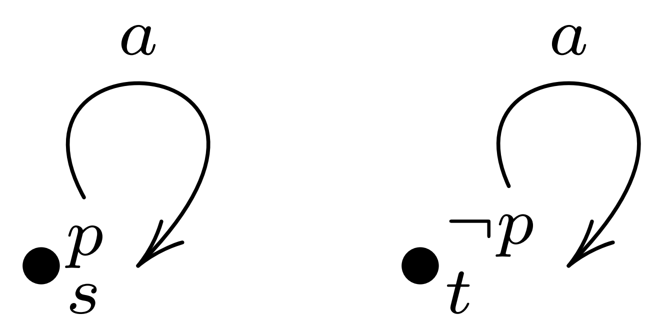
ϕ is a true lie by a in M, s iff $M, s \models B_a\neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$\phi_0 = p \wedge B_b p$

M_0:



$M_0|^a_{B_a\phi_0}$:



ϕ_0 is a true lie by a in M_0, s

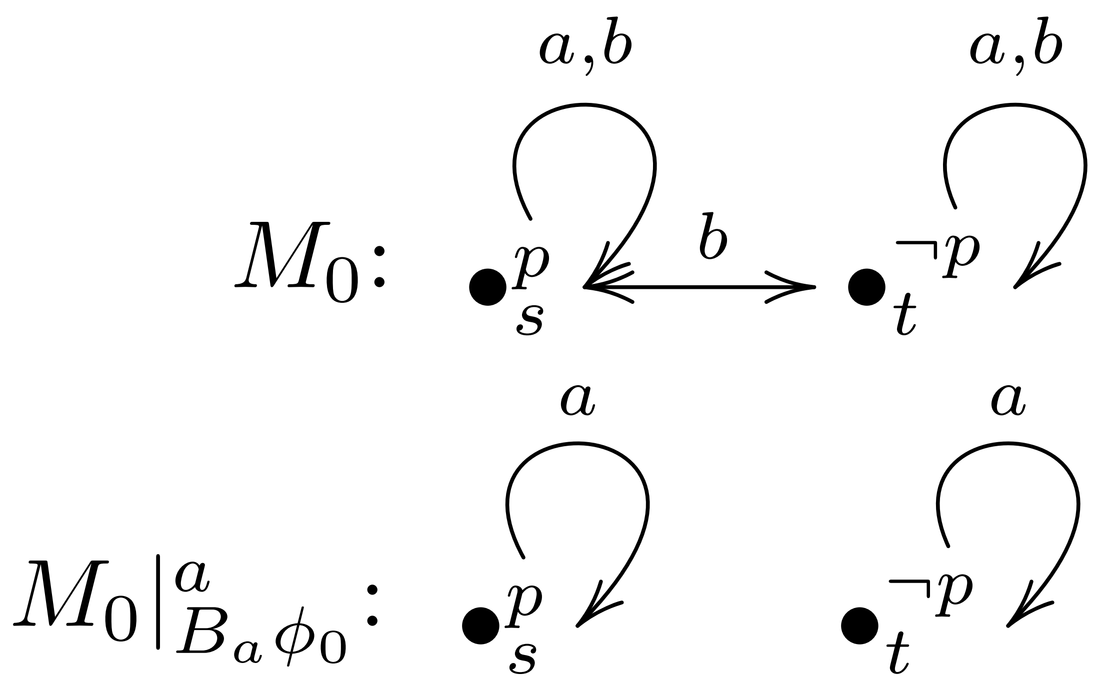
ϕ_0 is not a true lie by a in M_0, t

ϕ_0 is not a believable true lie by a in M_0, s

Example: from the inside

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$$\phi_0 = p \wedge B_b p$$



ϕ_0 is a true lie by a in M_0, s

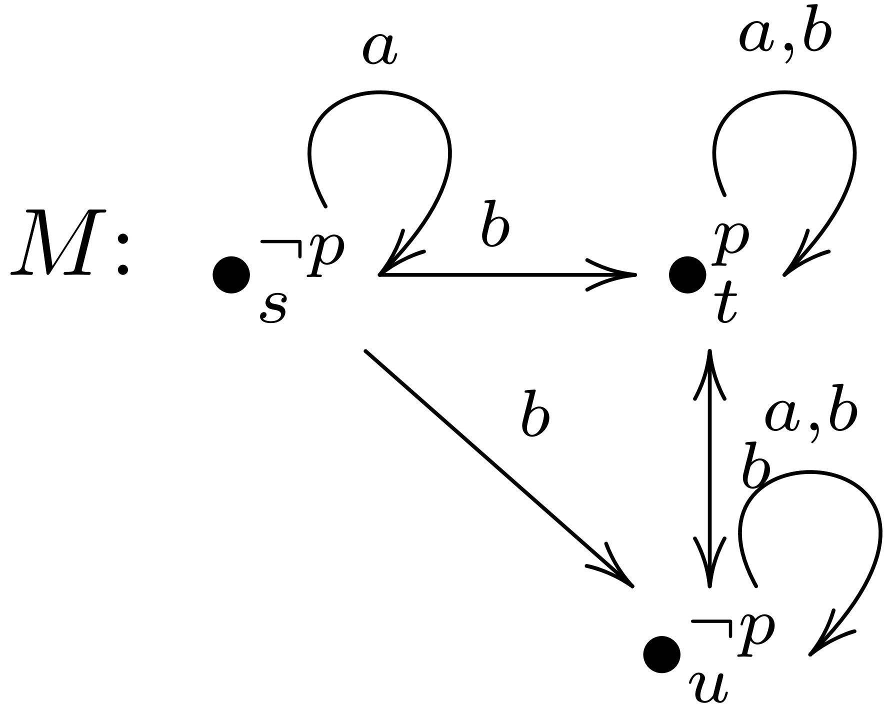
ϕ_0 is not a true lie by a in M_0, t

ϕ_0 is not a believable true lie by a in M_0, s
(it can be shown that ϕ_0 is not a believable true lie on any S5 model)

Example

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^{a}_{B_a\phi}, s \models \phi$

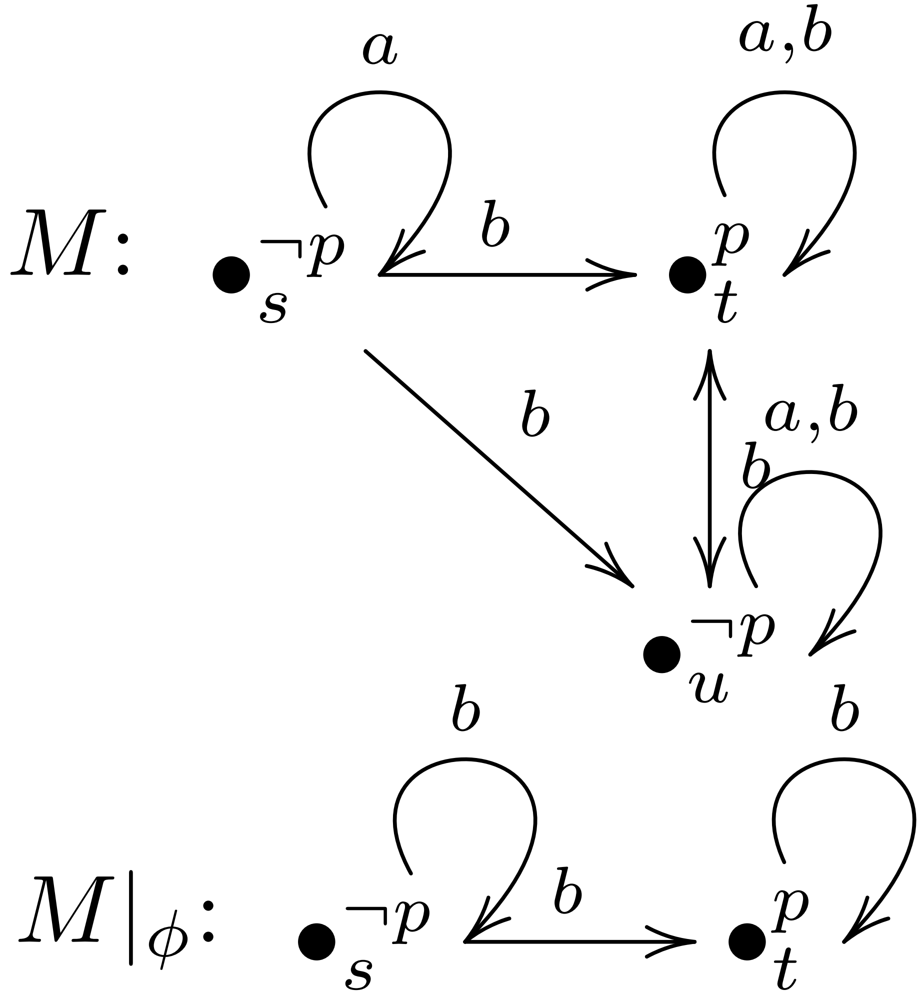
$$\phi = p \vee B_b p$$



Example

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg \phi$ and $M|^a_{B_a \phi}, s \models \phi$

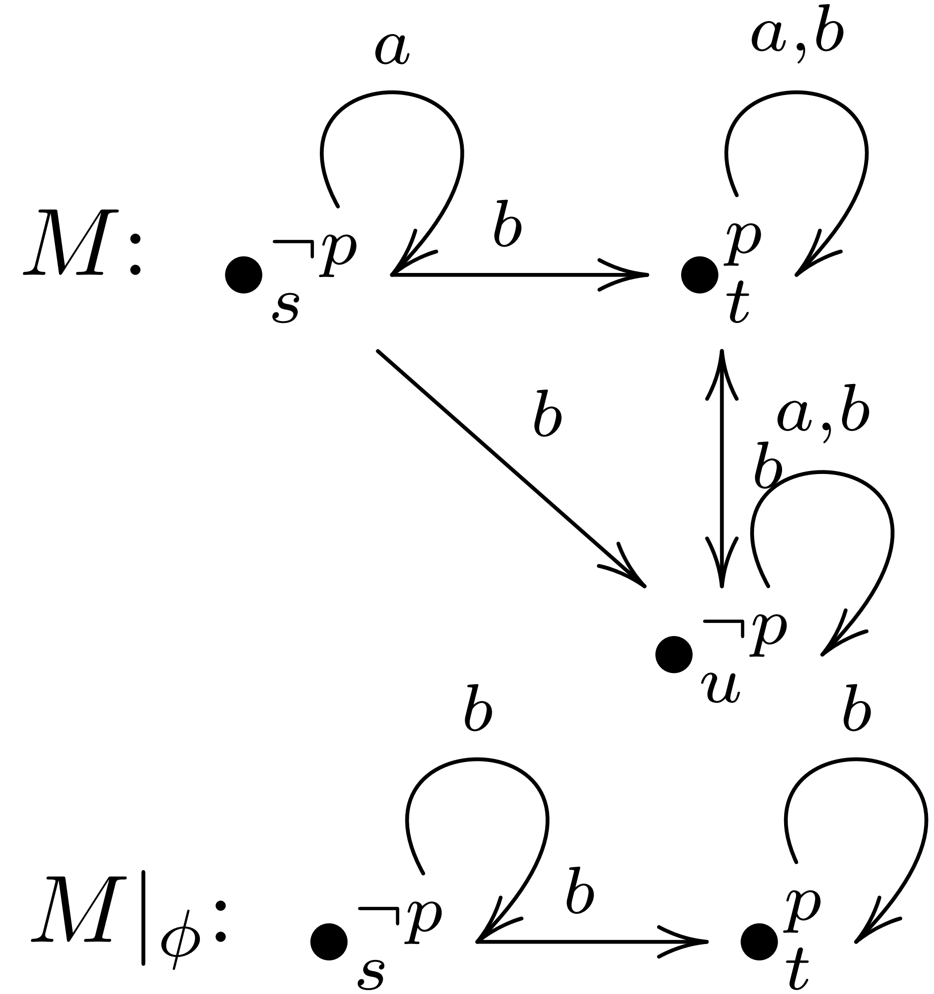
$$\phi = p \vee B_b p$$



Example

ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

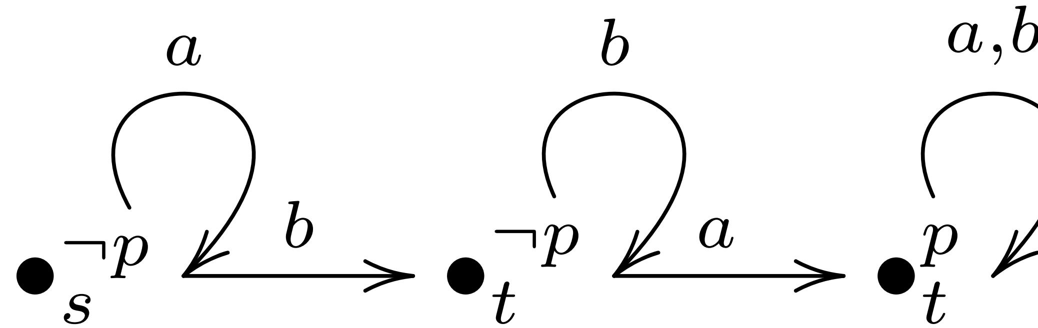
$$\phi = p \vee B_b p$$



ϕ is a believable true lie in M, s

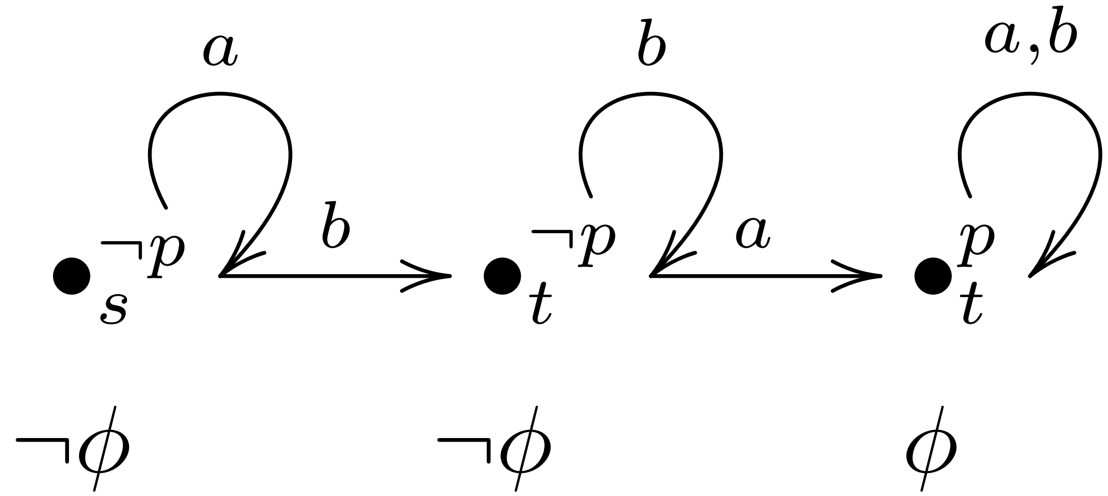
.. but not a proper true lie by a

$$\phi = p \vee B_b p$$



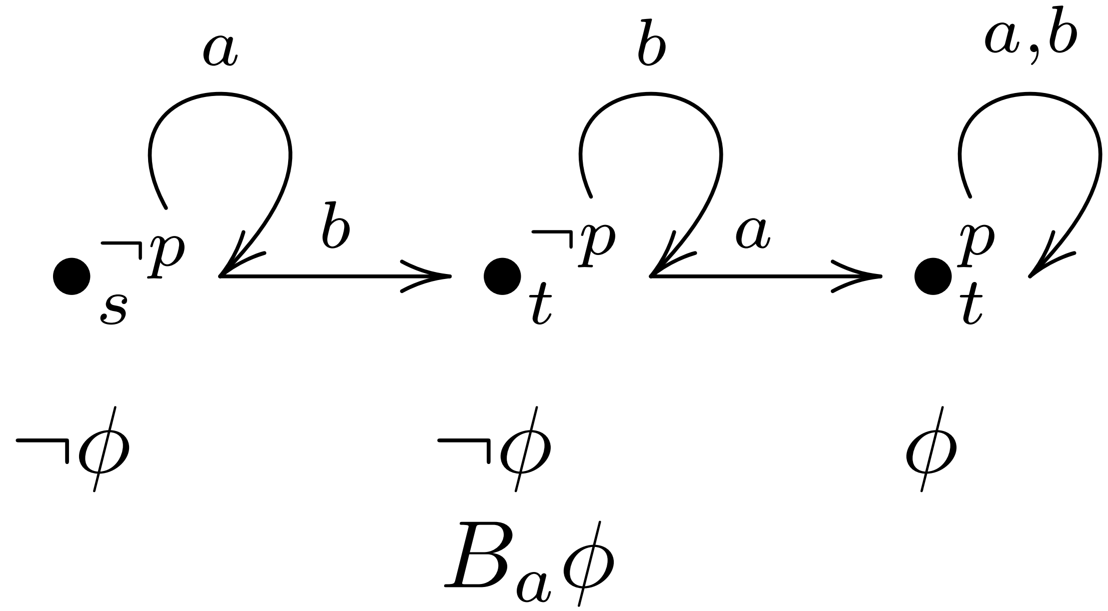
.. but not a proper true lie by a

$$\phi = p \vee B_b p$$



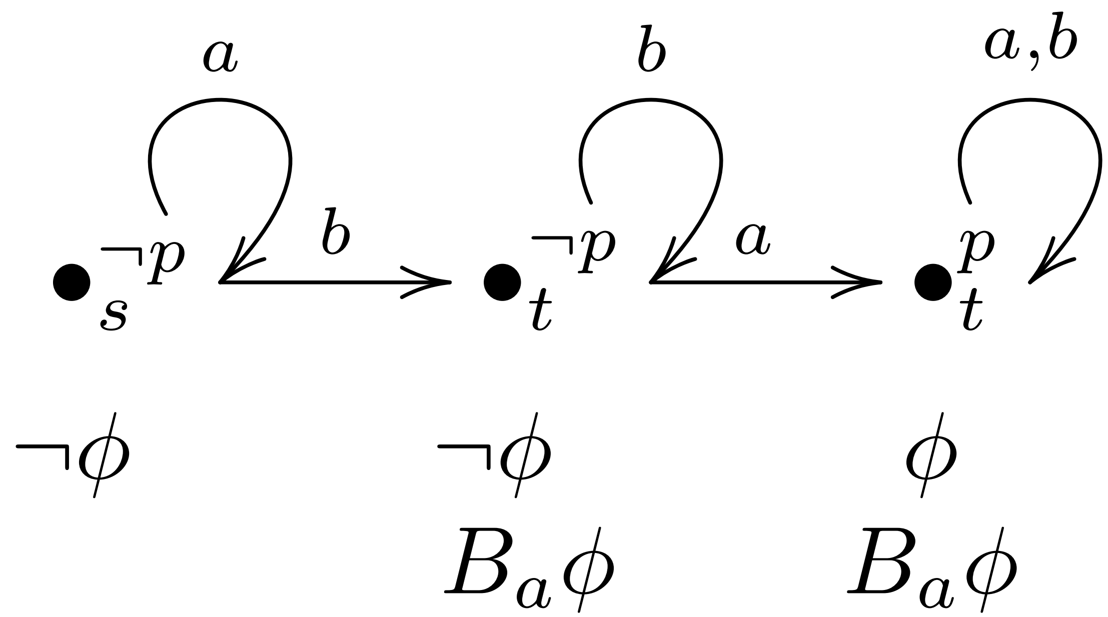
.. but not a proper true lie by a

$$\phi = p \vee B_b p$$



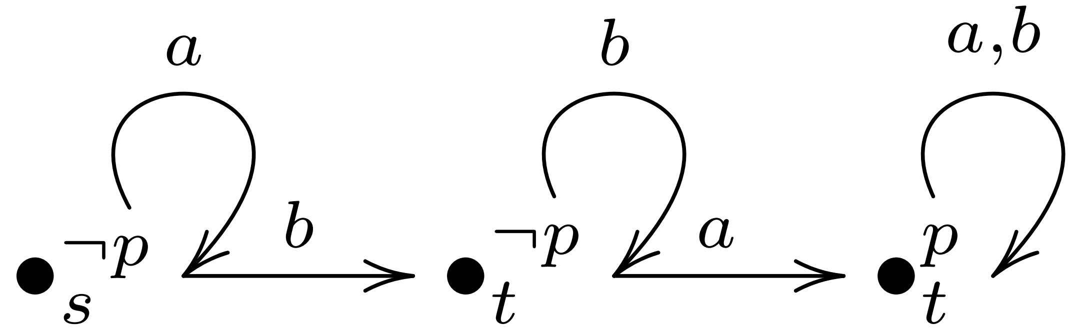
.. but not a proper true lie by a

$$\phi = p \vee B_b p$$



.. but not a proper true lie by a

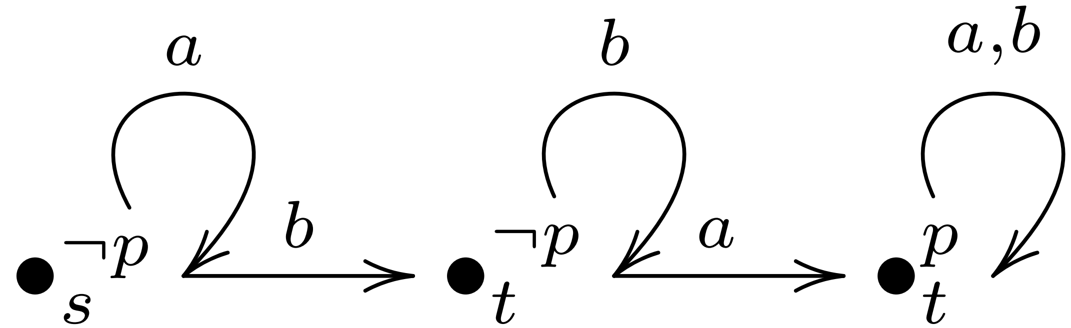
$$\phi = p \vee B_b p$$



$\neg\phi$	$\neg\phi$	ϕ
$\neg B_a\phi$	$B_a\phi$	$B_a\phi$

.. but not a proper true lie by a

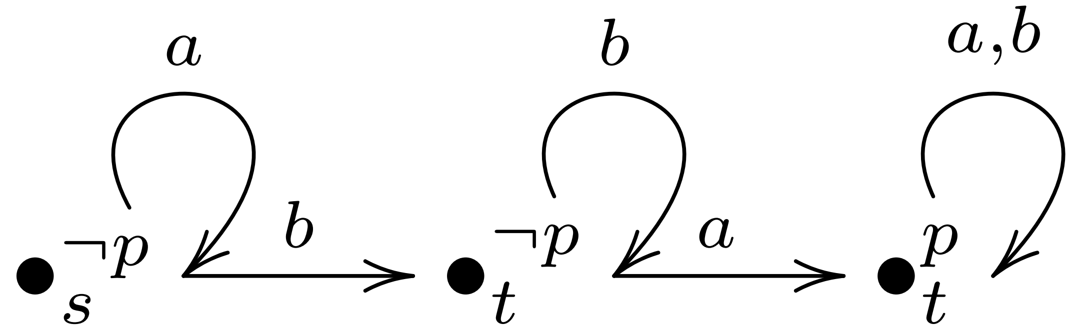
$$\phi = p \vee B_b p$$



$\neg\phi$	$\neg\phi$	ϕ
$\neg B_a \phi$	$B_a \phi$	$B_a \phi$
$B_a \neg\phi$		

.. but not a proper true lie by a

$$\phi = p \vee B_b p$$



$\neg\phi$	$\neg\phi$	ϕ
$\neg B_a\phi$	$B_a\phi$	$B_a\phi$
$B_a\neg\phi$		

ϕ is also not a believable true lie (same counterexample)

The Logic of Lying, and Private True Lies

Language with explicit public lies and announcements

$$\phi \ ::= \ \top \mid p \mid \neg\phi \mid \phi \wedge \phi \mid B\phi \mid \langle !\phi\rangle\phi \mid \langle ¡\phi\rangle\phi$$

$$\mathcal{U}_{¡\phi}: \quad \begin{tikzcd} \underline{\hat{B}\phi \wedge \neg\phi} \arrow[rr] & & \phi \arrow[loop above] \end{tikzcd}$$

Language with explicit public lies and announcements

$$\phi \;::=\; \top \mid p \mid \neg\phi \mid \phi \wedge \phi \mid B\phi \mid \langle !\phi \rangle \phi \mid \langle ¡\phi \rangle \phi$$

$$\mathcal{U}_{¡\phi} : \begin{tikzcd} \underline{\hat{B}\phi \wedge \neg\phi} \arrow[r] & \phi \arrow[loop above] \end{tikzcd}$$

TAUT	all the instances of tautologies
DISTK	$\star(\phi \to \psi) \to (\star\phi \to \star\psi)$
MP	$\dfrac{\phi, \phi \to \psi}{\psi}$
GEN	$\dfrac{\phi}{\star\phi}$
INV	$(p \to [!\psi]p) \wedge (\neg p \to [!\psi]\neg p)$
INV	$(p \to [¡\psi]p) \wedge (\neg p \to [¡\psi]\neg p)$
PRE	$\langle !\psi \rangle \top \leftrightarrow \psi$
PRE	$\langle ¡\psi \rangle \top \leftrightarrow (\neg\psi \wedge \hat{B}\psi)$
DET	$\langle ¡\psi \rangle \phi \to [¡\psi]\phi$
DET	$\langle !\psi \rangle \phi \to [¡\psi]\phi$
NM	$\langle ¡\psi \rangle B\phi \to B[!\psi]\phi$
NM	$\langle !\psi \rangle B\phi \to B[!\psi]\phi$
PR	$B[!\psi]\phi \to [¡\psi]B\phi$
PR	$B[!\psi]\phi \to [!\psi]B\phi$

Language with explicit public lies and announcements

$$\phi \ ::= \top \mid p \mid \neg\phi \mid \phi \wedge \phi \mid B\phi \mid \langle !\phi\rangle\phi \mid \langle \text{¡}\phi\rangle\phi$$

TAUT	all the instances of tautologies
DISTK	$\star(\phi \to \psi) \to (\star\phi \to \star\psi)$
MP	$\dfrac{\phi, \phi \to \psi}{\psi}$
GEN	$\dfrac{\phi}{\star\phi}$
INV	$(p \to [!\psi]p) \wedge (\neg p \to [!\psi]\neg p)$
INV	$(p \to [\text{¡}\psi]p) \wedge (\neg p \to [\text{¡}\psi]\neg p)$
PRE	$\langle !\psi\rangle\top \leftrightarrow \psi$
PRE	$\langle \text{¡}\psi\rangle\top \leftrightarrow (\neg\psi \wedge \hat{B}\psi)$
DET	$\langle \text{¡}\psi\rangle\phi \to [\text{¡}\psi]\phi$
DET	$\langle !\psi\rangle\phi \to [\text{¡}\psi]\phi$
NM	$\langle \text{¡}\psi\rangle B\phi \to B[!\psi]\phi$
NM	$\langle !\psi\rangle B\phi \to B[!\psi]\phi$
PR	$B[!\psi]\phi \to [\text{¡}\psi]B\phi$
PR	$B[!\psi]\phi \to [!\psi]B\phi$

$$\mathcal{U}_{\text{¡}\phi} : \begin{tikzcd} \underline{\hat{B}\phi \wedge \neg\phi} \arrow[rr] & & \phi \arrow[loop above] \end{tikzcd}$$

Sound and complete

Language with implicit public lies and announcements

$$\phi \ ::= \ \top \mid p \mid \neg\phi \mid \phi \wedge \phi \mid B\phi \mid \langle ¡\phi \rangle \phi$$

TAUT	all the instances of tautologies
DISTK	$\star(\phi \to \psi) \to (\star\phi \to \star\psi)$
MP	$\dfrac{\phi, \phi \to \psi}{\psi}$
GEN	$\dfrac{\phi}{\star\phi}$
INV	$(p \to [¡\psi]p) \wedge (\neg p \to [¡\psi]\neg p)$
PRE	$\langle ¡\psi \rangle \top \leftrightarrow (\psi \vee (\hat{B}\psi \wedge \neg\psi))$
DET	$\langle ¡\psi \rangle \phi \to [¡\psi]\phi$
NM	$\langle ¡\psi \rangle B\phi \to B(\psi \to [¡\psi]\phi)$
PR	$B(\psi \to [¡\psi]\phi) \to [¡\psi]B\phi$

$$\mathcal{M}, w \vDash \langle ¡\psi \rangle \phi \iff \mathcal{M}, w \vDash \psi \vee (\neg\psi \wedge \hat{B}\psi) \text{ and } \mathcal{M} \otimes \mathcal{U}, (w, u) \vDash \phi$$

where u is the ϕ-action iff $\mathcal{M}, w \vDash \psi$.

$$\begin{tikzcd} \mathcal{U}_{¡\psi} : \ \underline{\hat{B}\psi \wedge \neg\psi} \arrow[rr] & & \underline{\psi} \arrow[loop above] \end{tikzcd}$$

Language with explicit private lies and announcements

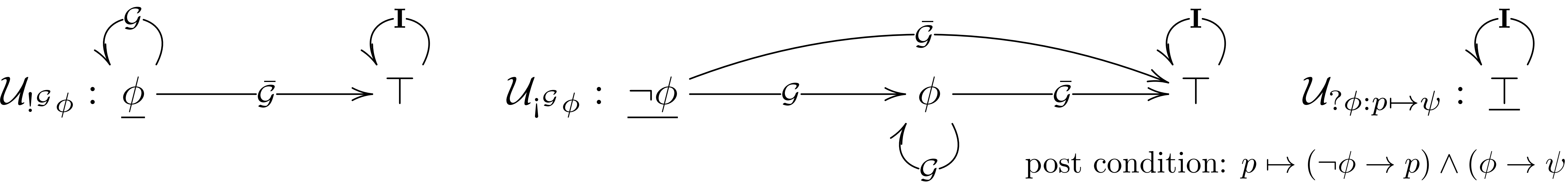
$$\phi \ ::= \ \top \mid p \mid \neg\phi \mid \phi \wedge \phi \mid B_i\phi \mid \langle !^{\mathcal{G}}\phi\rangle\phi \mid \langle ¡^{\mathcal{G}}\phi\rangle\phi \mid \langle ?\phi : p \mapsto \phi\rangle\phi$$

$$\begin{tikzcd}
\mathcal{U}_{!^{\mathcal{G}}\phi} : & \underline{\phi} \arrow[loop above, "\mathcal{G}"] \arrow[r, "\bar{\mathcal{G}}"] & \top \arrow[loop above, "\mathbf{I}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathcal{U}_{¡^{\mathcal{G}}\phi} : & \underline{\neg\phi} \arrow[r, "\mathcal{G}"] \arrow[rr, bend left, "\bar{\mathcal{G}}"] & \phi \arrow[loop below, "\mathcal{G}"] \arrow[r, "\bar{\mathcal{G}}"] & \top \arrow[loop above, "\mathbf{I}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathcal{U}_{?\phi:p\mapsto\psi} : & \underline{\top} \arrow[loop above, "\mathbf{I}"]
\end{tikzcd}$$

post condition: $p \mapsto (\neg\phi \to p) \wedge (\phi \to \psi)$

Language with explicit private lies and announcements

$$\phi \ ::= \ \top \mid p \mid \neg\phi \mid \phi\wedge\phi \mid B_i\phi \mid \langle !^{\mathcal{G}}\phi\rangle\phi \mid \langle ¡^{\mathcal{G}}\phi\rangle\phi \mid \langle ?\phi : p\mapsto\phi\rangle\phi$$



TAUT	all the instances of tautologies
DISTK	$\star(\phi\to\psi)\to(\star\phi\to\star\psi)$
MP	$\dfrac{\phi,\phi\to\psi}{\psi}$
GEN	$\dfrac{\phi}{\star\phi}$
INV	$(p\to[!^{\mathcal{G}}\phi]p)\wedge(\neg p\to[!^{\mathcal{G}}\phi]\neg p)$
INV	$(p\to[¡^{\mathcal{G}}\phi]p)\wedge(\neg p\to[¡^{\mathcal{G}}\phi]\neg p)$
INV	$((\neg\phi\to p)\wedge(\phi\to\psi))\leftrightarrow[?\phi:p\mapsto\psi]p$
PRE	$\langle !^{\mathcal{G}}\phi\rangle\top\leftrightarrow\phi$
PRE	$\langle ¡^{\mathcal{G}}\phi\rangle\top\leftrightarrow\neg\phi$
PRE	$\langle ?\phi:p\mapsto\psi\rangle\top$

DET	$\langle ¡^{\mathcal{G}}\phi\rangle\phi\to[¡^{\mathcal{G}}\phi]\phi$
DET	$\langle !^{\mathcal{G}}\phi\rangle\phi\to[¡^{\mathcal{G}}\phi]\phi$
DET	$\langle ?\phi:p\mapsto\psi\rangle\chi\to[?\phi:p\mapsto\psi]\chi$
UB$(i\in\mathcal{G})$	$[¡^{\mathcal{G}}\psi]B_i\phi\leftrightarrow(\neg\psi\to B_i[!^{\mathcal{G}}\psi]\phi)$
UB$(i\notin\mathcal{G})$	$[¡^{\mathcal{G}}\psi]B_i\phi\leftrightarrow(\neg\psi\to B_i\phi)$
UB$(i\in\mathcal{G})$	$[!^{\mathcal{G}}\psi]B_i\phi\leftrightarrow(\psi\to B_i[!^{\mathcal{G}}\psi]\phi)$
UB$(i\notin\mathcal{G})$	$[!^{\mathcal{G}}\psi]B_i\phi\leftrightarrow(\psi\to B_i\phi)$
UB	$[?\phi:p\mapsto\psi]B_i\phi\leftrightarrow(B_i[?\phi:p\mapsto\psi]\phi)$

The party example

$$
\begin{tikzcd}
\underline{\neg p_1, \neg p_2} \arrow[r, dash, "1"] \arrow[d, dash, "2"] & \neg p_1 p_2 \arrow[d, dash, "2"] \\
p_1, \neg p_2 \arrow[r, dash, "1"] & p_1, p_2
\end{tikzcd}
$$

The party example

$$
\begin{tikzcd}
\underline{\neg p_1, \neg p_2} \arrow[r, dash, "1"] \arrow[d, dash, "2"] & \neg p_1 p_2 \arrow[d, dash, "2"] \\
p_1, \neg p_2 \arrow[r, dash, "1"] & p_1, p_2
\end{tikzcd}
$$

The update model $\mathcal{U}$ for $\text{¡}^1 p_2$:

$$
\begin{tikzcd}
\mathcal{U}_{\text{¡}^1 p_2} : \ \underline{\neg p_2} \arrow[r, "1"] \arrow[rr, bend left, "2"] & p_2 \arrow[r, "2"] \arrow[loop below, "1"] & \top \arrow[loop above, "1,2"]
\end{tikzcd}
$$

The party example

$$\begin{tikzcd}
\underline{\neg p_1 , \neg p_2} \arrow[r, dash, "1"] \arrow[d, dash, "2"] & \neg p_1 p_2 \arrow[d, dash, "2"] \\
p_1 , \neg p_2 \arrow[r, dash, "1"] & p_1 , p_2
\end{tikzcd}$$

The update model $\mathcal{U}$ for $\text{¡}^1 p_2$:

$$\begin{tikzcd}
\mathcal{U}_{\text{¡}^1 p_2} : \; \underline{\neg p_2} \arrow[r, "1"] \arrow[rr, bend left, "2"] & p_2 \arrow[r, "2"] \arrow[loop below, "1"] & \top \arrow[loop above, "1,2"]
\end{tikzcd}$$

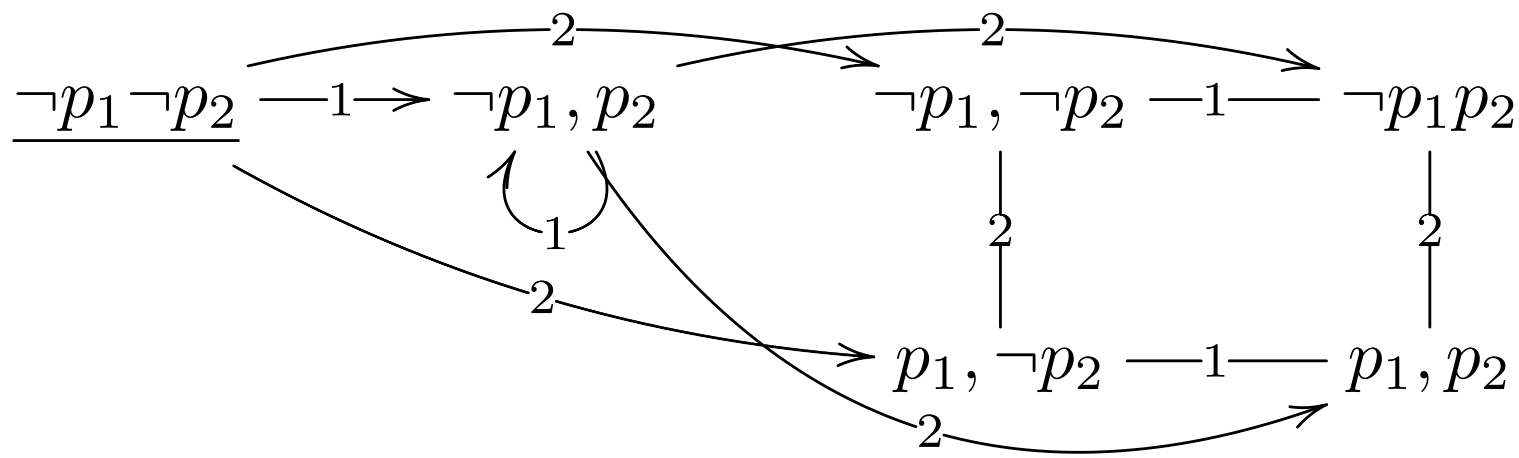
Updated model ($\mathcal{M} \otimes \mathcal{U}$)

$$\begin{tikzcd}
\underline{\neg p_1 \neg p_2} \arrow[r, "1"] \arrow[rr, bend left, "2"] \arrow[drr, bend right, "2"] & \neg p_1 , p_2 \arrow[loop below, "1"] \arrow[rr, bend left, "2"] \arrow[drr, bend right, "2"] & \neg p_1 , \neg p_2 \arrow[r, dash, "1"] \arrow[d, dash, "2"] & \neg p_1 p_2 \arrow[d, dash, "2"] \\
 & & p_1 , \neg p_2 \arrow[r, dash, "1"] & p_1 , p_2
\end{tikzcd}$$

Example (continued)

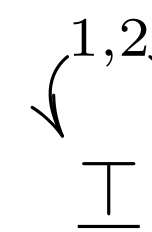
Example (continued)

Updated model ($\mathcal{M} \otimes \mathcal{U}$)

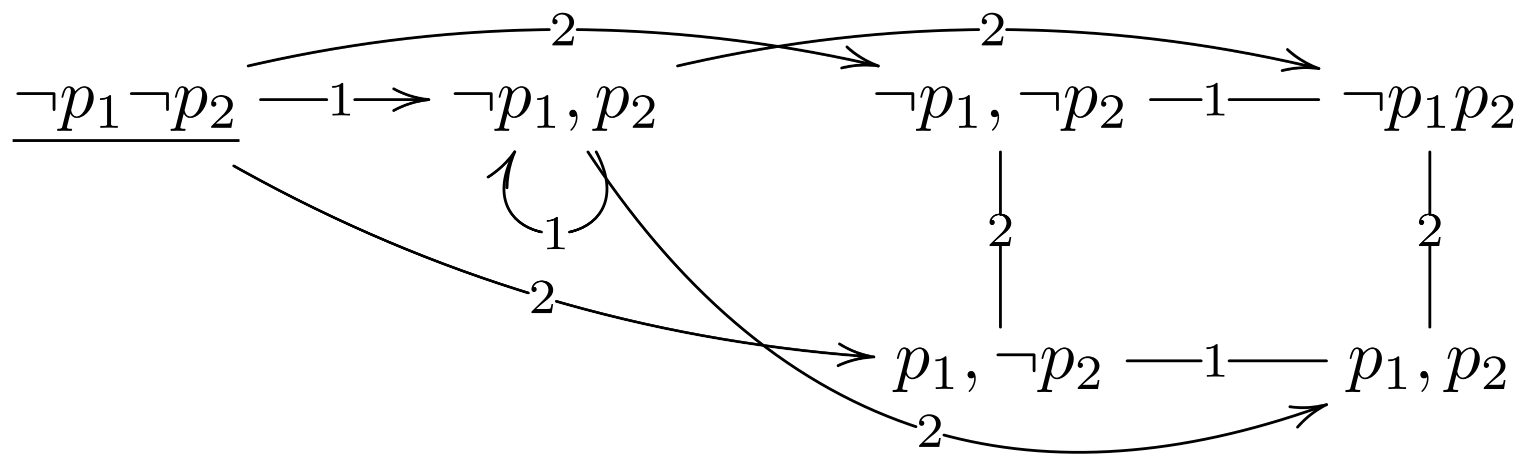


Example (continued)

The update model $\mathcal{U}'$ for $?B_1 p_2 : p_1 \mapsto \top$:

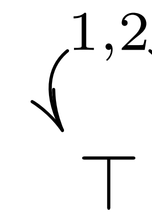


Updated model ($\mathcal{M} \otimes \mathcal{U}$)

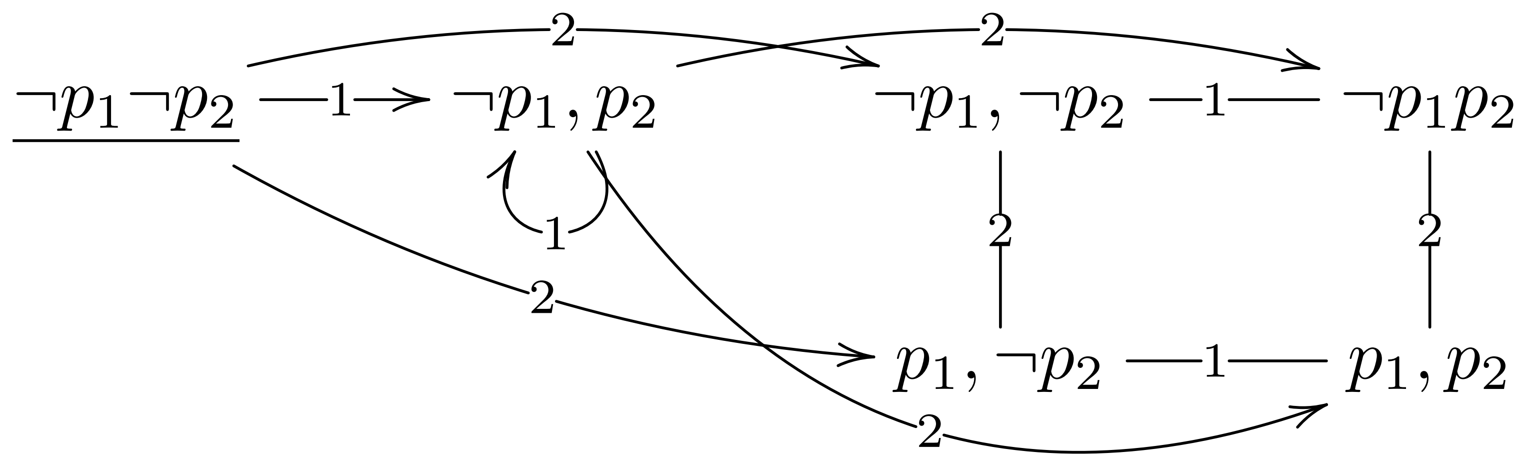


Example (continued)

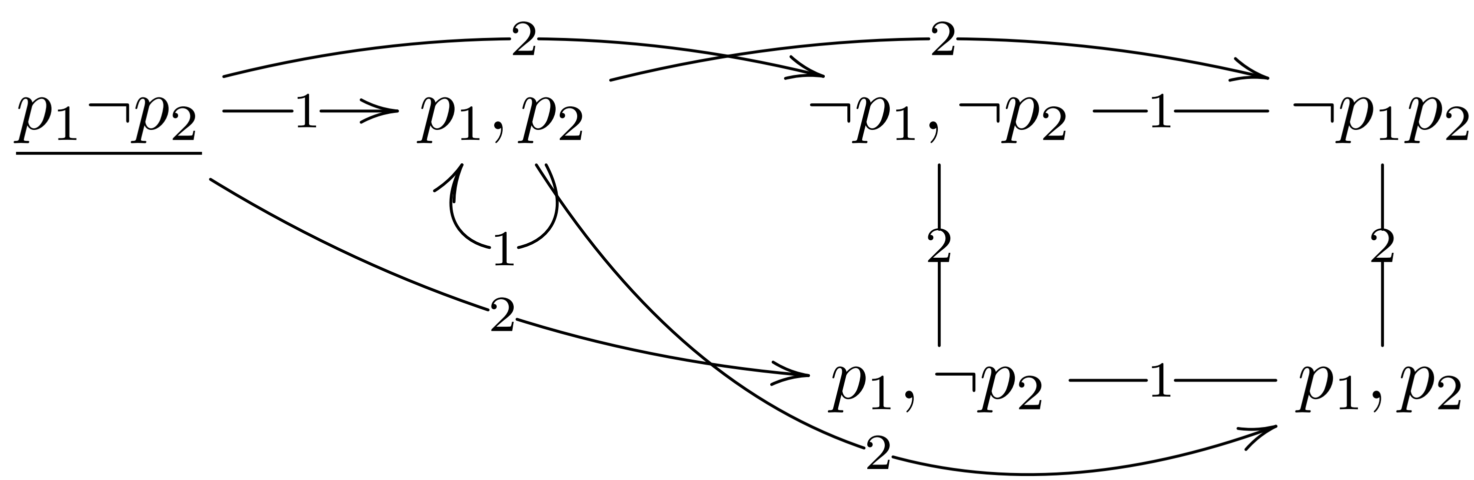
The update model $\mathcal{U}'$ for $?B_1 p_2 : p_1 \mapsto \top$:



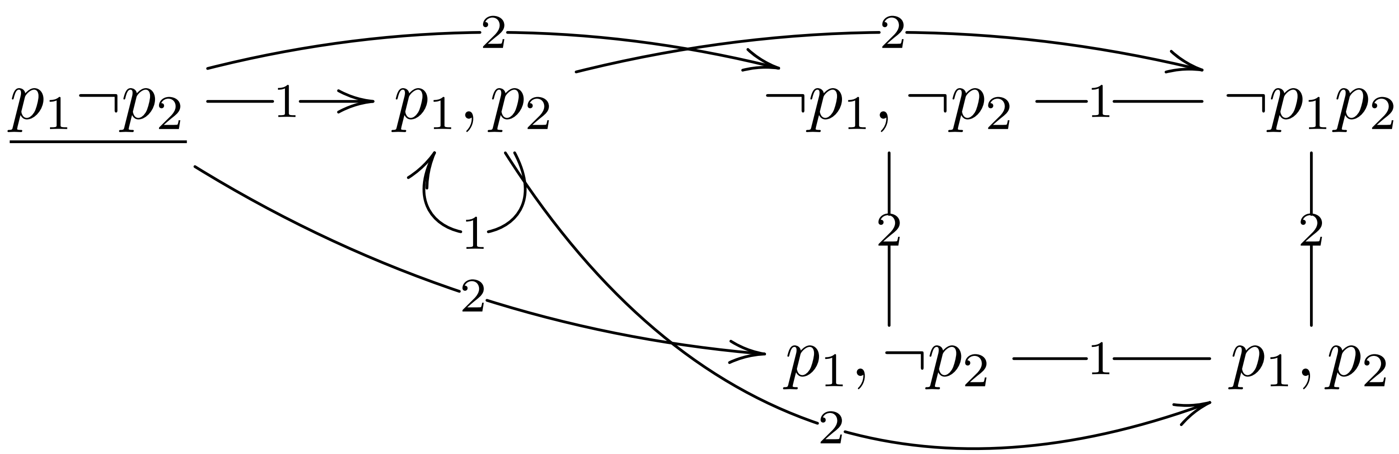
Updated model $(\mathcal{M} \otimes \mathcal{U})$



Updated model $(\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}')$

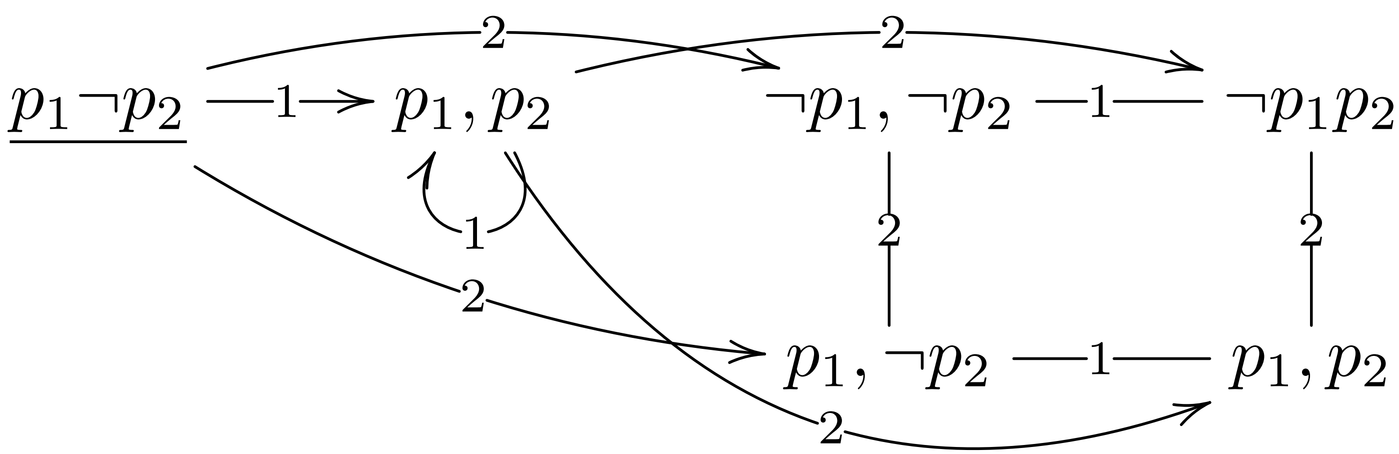


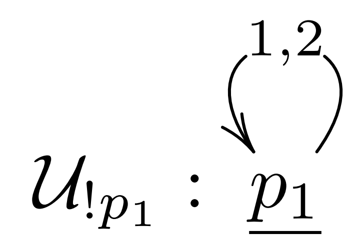
Example (continued)



Example (continued)

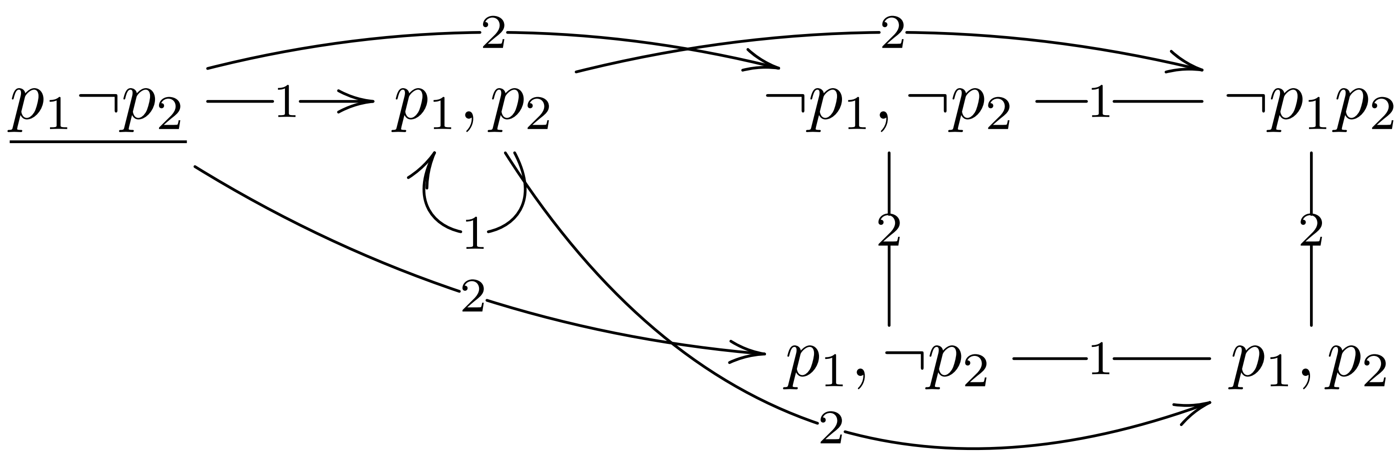
The update model $\mathcal{U}''$ for $!p_1$:

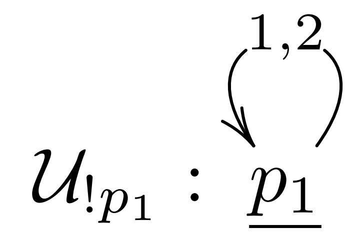




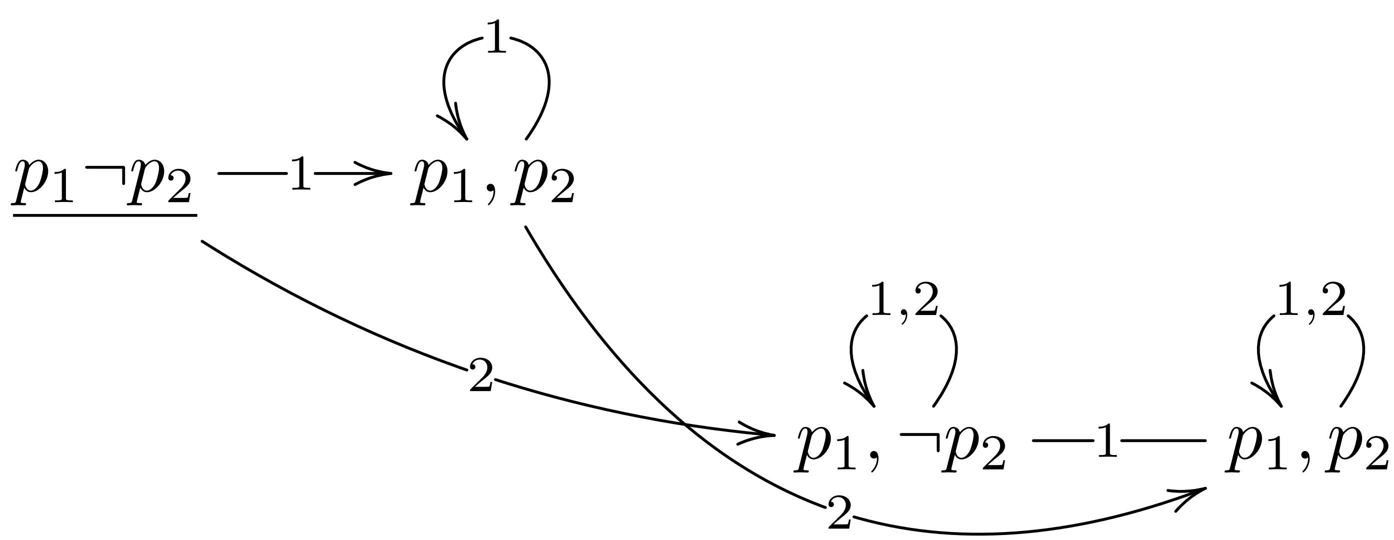
Example (continued)

The update model $\mathcal{U}''$ for $!p_1$:

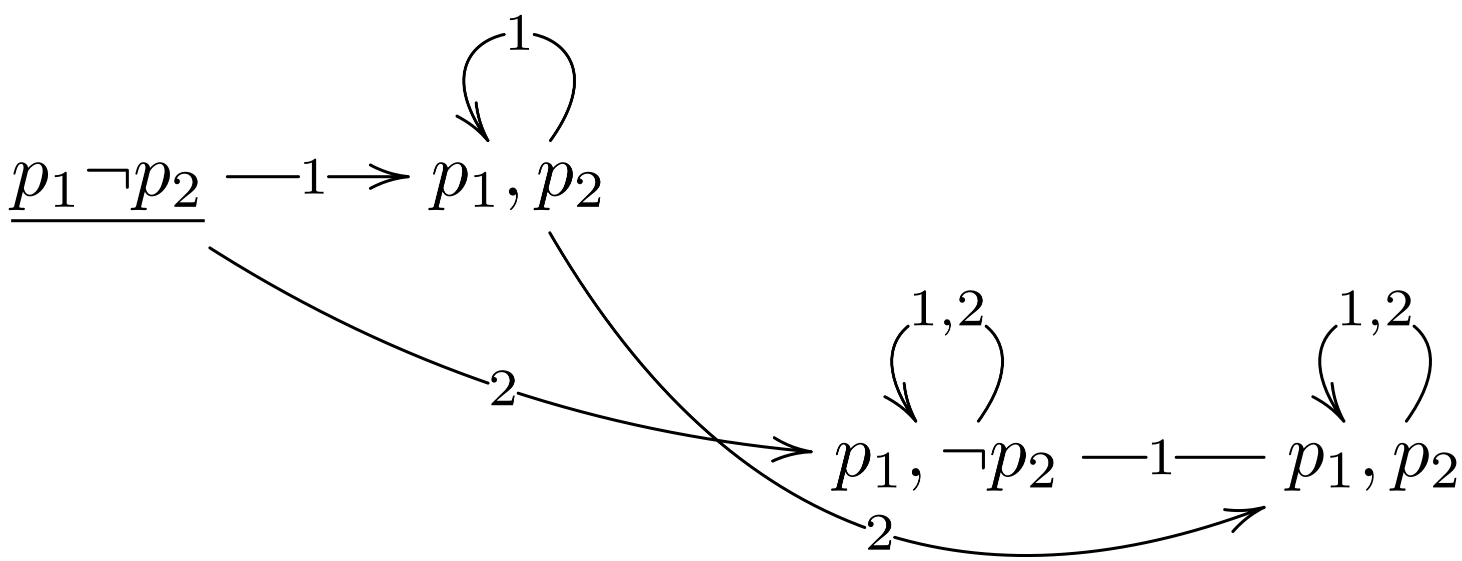




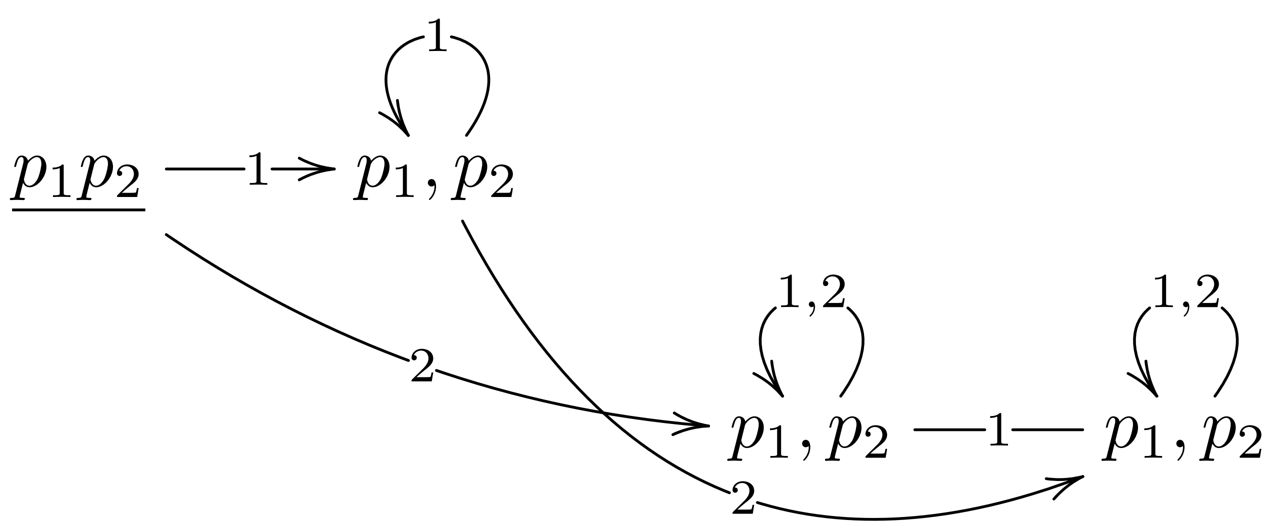
Updated model ($\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}' \otimes \mathcal{U}''$)



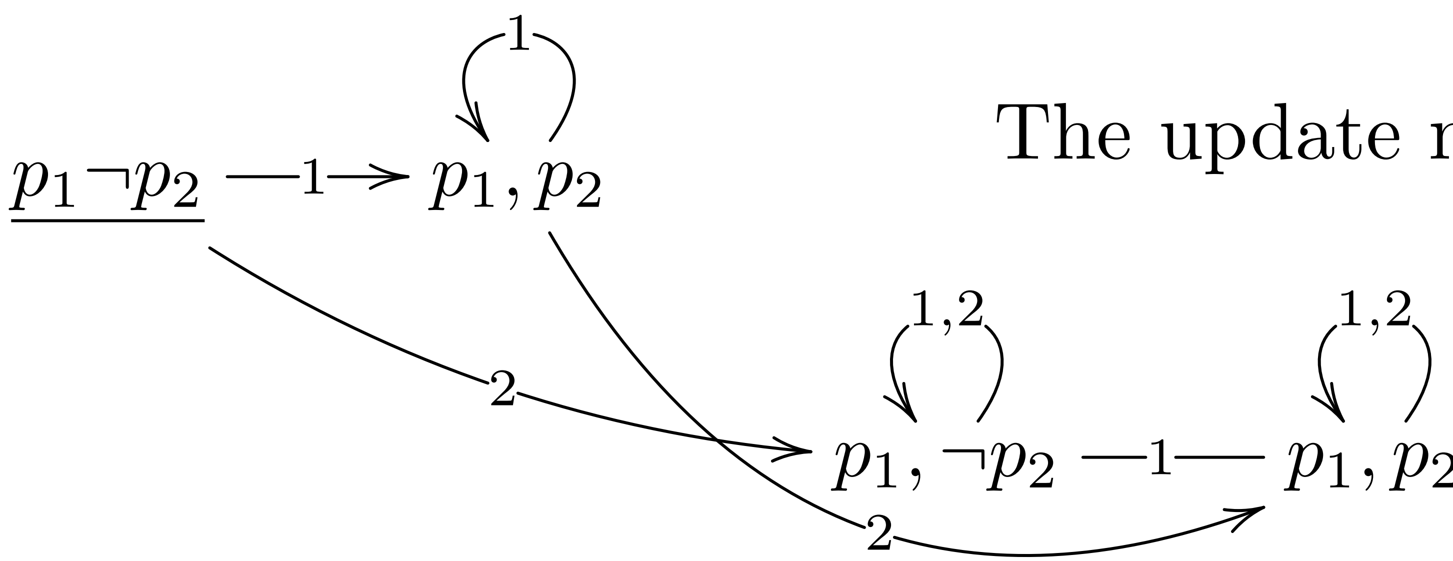
Example (continued)



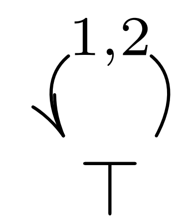
Updated model $(\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}' \otimes \mathcal{U}'' \otimes \mathcal{U}''')$



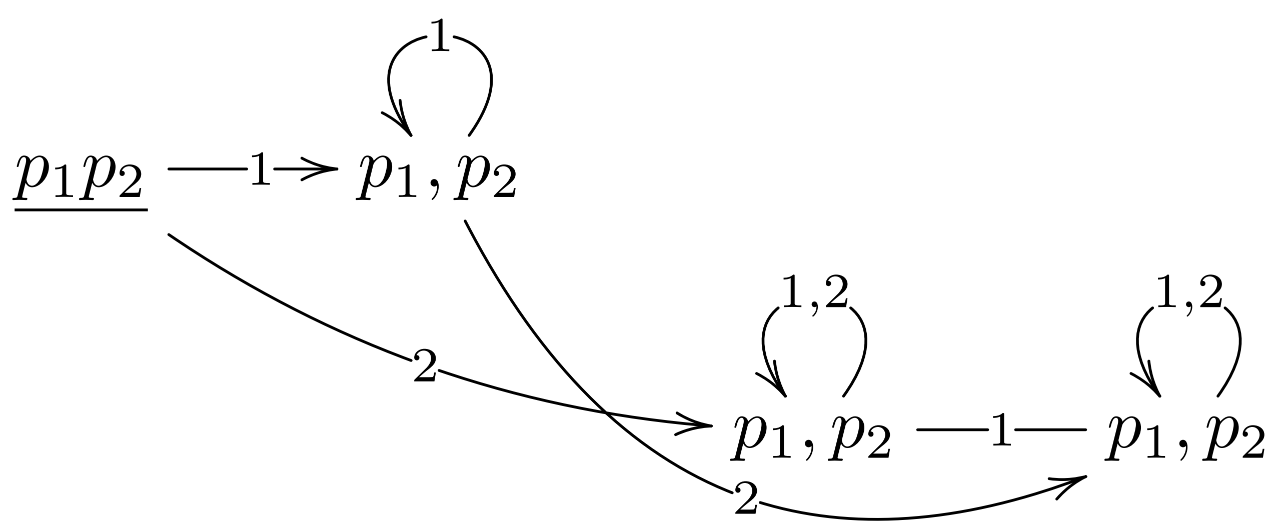
Example (continued)



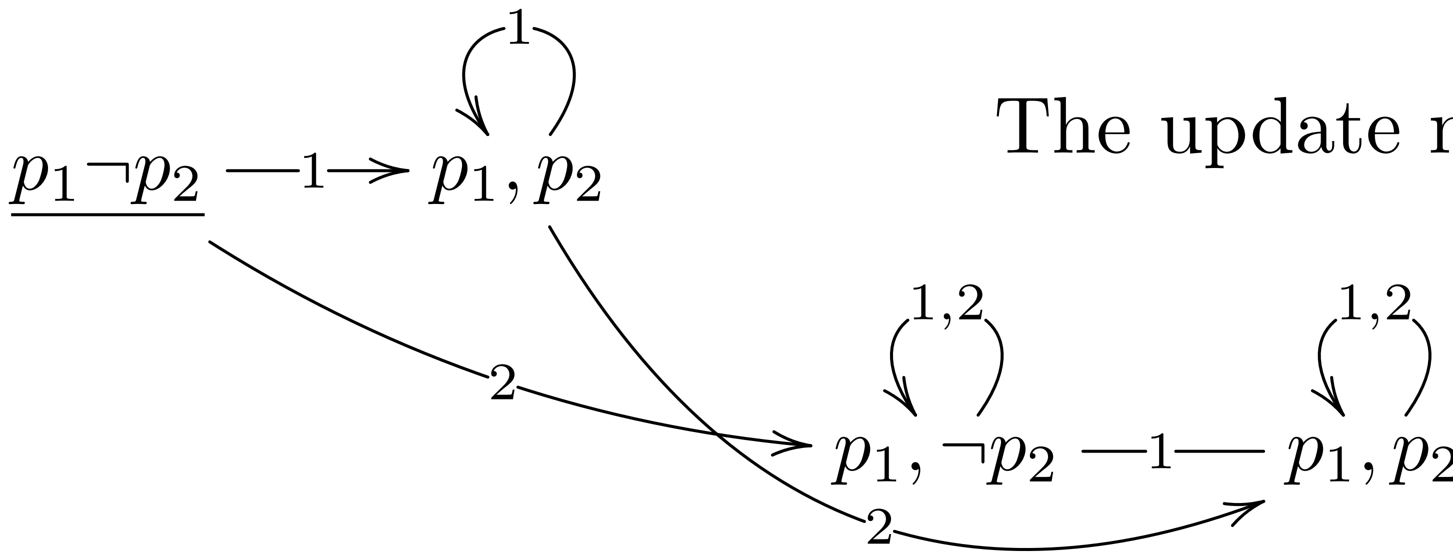
The update model $\mathcal{U}'''$ for $?B_2p_1 : p_2 \mapsto \top$:



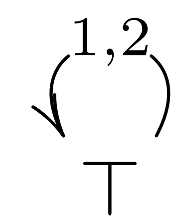
Updated model ($\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}' \otimes \mathcal{U}'' \otimes \mathcal{U}'''$)



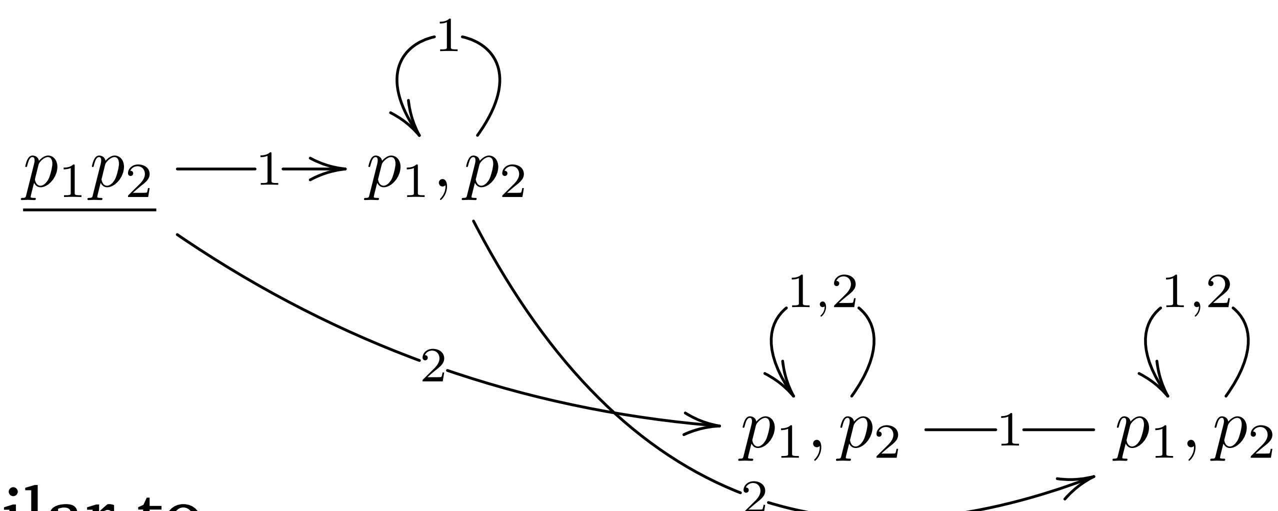
Example (continued)



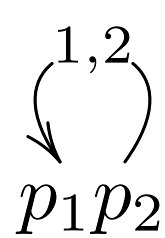
The update model $\mathcal{U}'''$ for $?B_2 p_1 : p_2 \mapsto \top$:



Updated model ($\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}' \otimes \mathcal{U}'' \otimes \mathcal{U}'''$)



which is similar to



Example (continued)

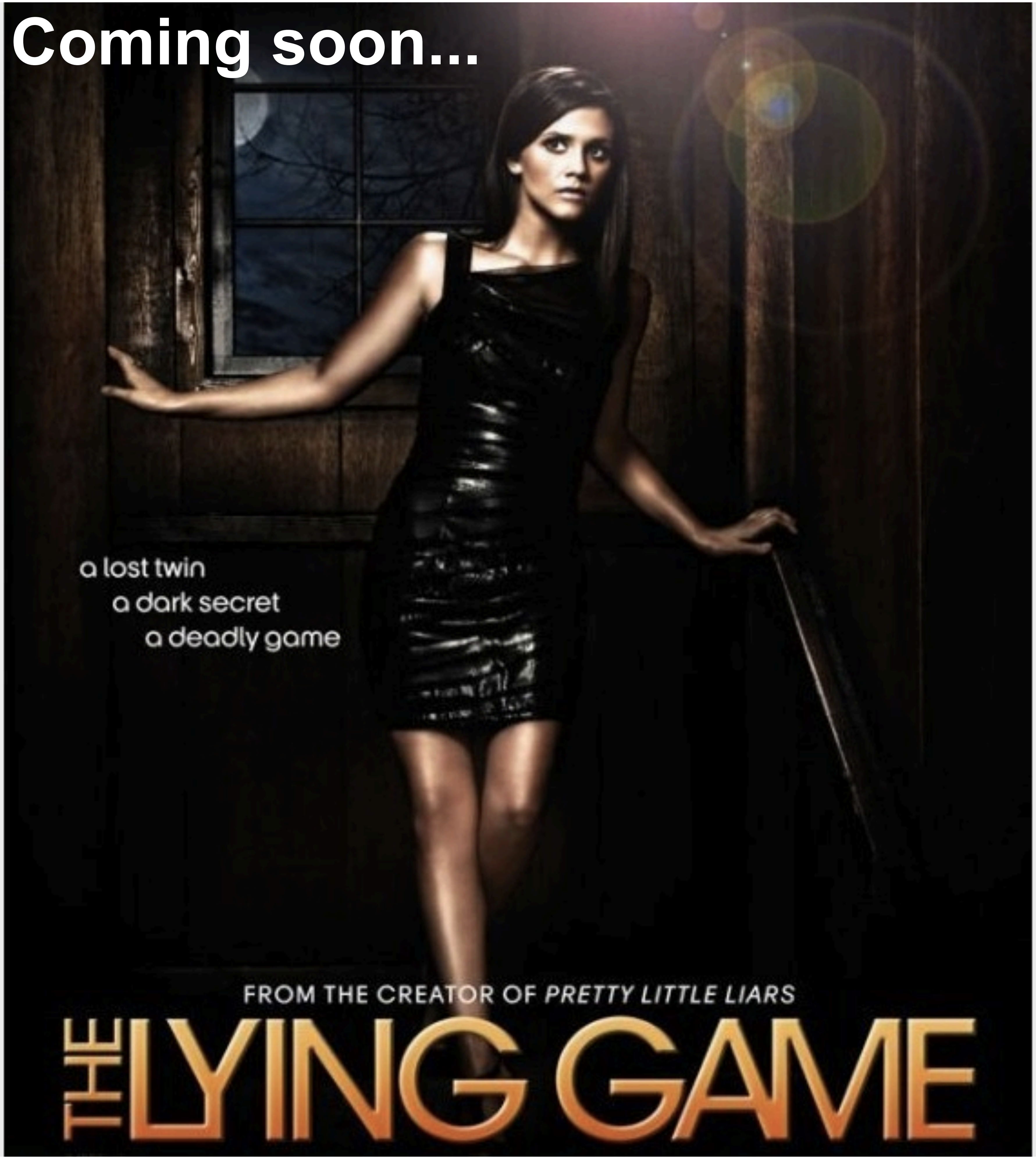
$$\begin{tikzcd}
\underline{\neg p_1, \neg p_2} \arrow[r, dash, "1"] \arrow[d, dash, "2"] & \neg p_1 p_2 \arrow[d, dash, "2"] \\
p_1, \neg p_2 \arrow[r, dash, "1"] & p_1, p_2
\end{tikzcd}$$

$$\mathcal{M}, w \vDash \neg p_1 \wedge \neg p_2 \wedge \langle ¡_1 p_2 \rangle \langle ?B_1 p_2 : p_1 \mapsto \top \rangle \langle ! p_1 \rangle \langle ?B_2 p_1 : p_2 \mapsto \top \rangle p_1 \wedge p_2 \wedge B_{1,2}(p_1 \wedge p_2)$$

Summary

- Motivation
  - formalising true lies
  - understanding certain monotonicity properties of public announcement logic
- Related to other Moorean phenomena
- Future work:
  - Characterisation: the multi-agent case
  - Alternation questions
  - Understanding relationships
  - Lying games

Coming soon...


a lost twin
a dark secret
a deadly game
FROM THE CREATOR OF PRETTY LITTLE LIARS
THE LYING GAME