True Lies

SOCREAL, Sapporo, 25 October 2013

Thomas Ågotnes
University of Bergen, Norway, and
Southwest University, Chongqing, China

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
  - Formalising true lies
  - The logic of true lies

Background

Public Announcement Logic (Plaza, 1989)

$$\varphi ::= p \mid K_i\varphi \mid \neg\varphi \mid \varphi_1 \wedge \varphi_2 \mid \langle\varphi_1\rangle\varphi_2$$

ϕ_1 is true, and ϕ_2 is true after ϕ_1 is announced

Formally:

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

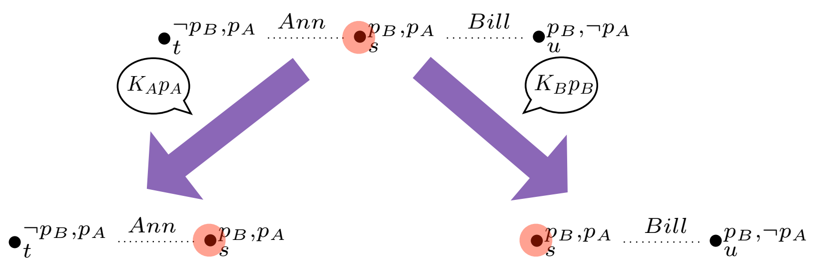
$$M, s \models K_i\phi \quad \Leftrightarrow \quad \forall t \sim_i s\ M, t \models \phi$$

$$M, s \models \langle\phi_1\rangle\phi_2 \quad \Leftrightarrow \quad M, s \models \phi_1 \text{ and } M|\phi_1, s \models \phi_2$$

The model resulting from removing states where ϕ_1 is false

Dual: $\hat{K}_i\phi \equiv \neg K_i\neg\phi$

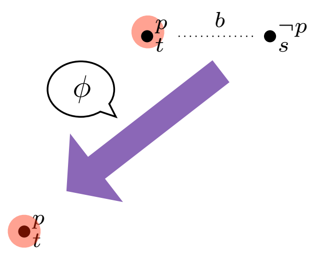
Example



$M, s \models \langle K_A p_A \rangle K_B p_A$

$M, s \models \langle K_B p_B \rangle K_A p_B$

Moore sentences



$\phi = p \wedge \neg K_b p$

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

Formalising true lies

Lies

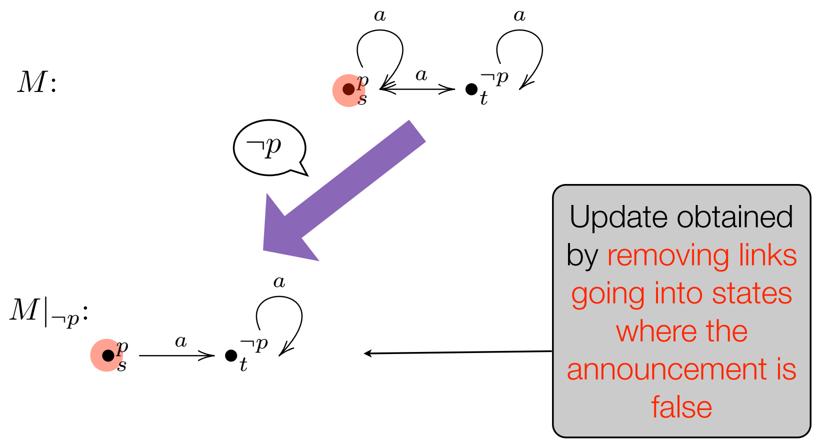
- Dimensions:
  - Who is the lier: one of the agents in the system, or an outsider?
  - Who are being lied to (and what do the others know about that)?
  - What are the agent's attitude to possible lies?
    - Credulous agents: believe everything
    - Skeptical agents: believe everything consistent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

Lies

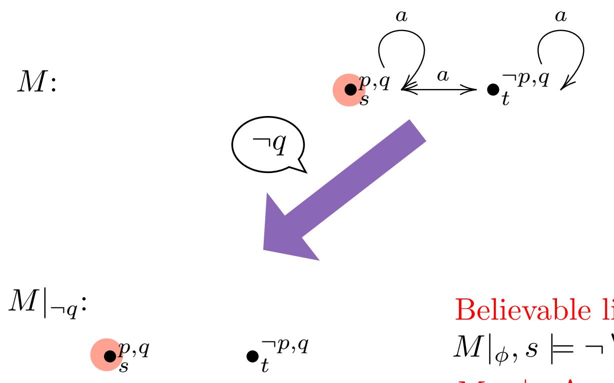
- Here:
  - Two cases: one of the agents in the system + outside observer
  - Public lie, to all other agents
  - Credulous/skeptical agents

True lies from the outside

Untruthful announcements: link-cutting semantics



Unbelievable lie



Believable lie:

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

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

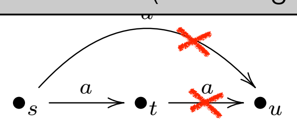
Models of lying

Already seen:

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

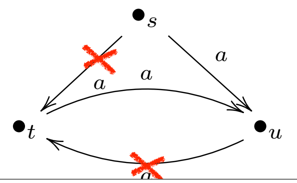
We will write B (belief) instead of K (knowledge)

Preservation of transitivity:



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

Prese city:



Formalising lies: made by an agent outside the system

Given: pointed model M, s

Pre-condition: $M, s \models \neg\phi$

Additional pre-condition for believable lies:
$M, s \models \bigwedge_{i \in Ag} \hat{B}_i \phi$

Consequence: $M|_\phi, s$ obtained by cutting links to $\neg\phi$-states for all agents

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 $\forall M \forall s : (M, s \models \neg\phi \Rightarrow M|_\phi, s \models \phi$

believable

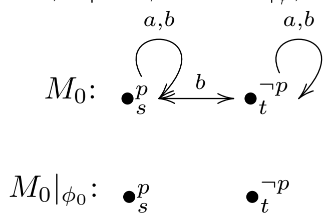
and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \bot)$

In some model class (typically K(D)45)

Example: from the outside

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

$\phi_0 = p \wedge B_b p$



ϕ_0 is a true lie in M_0, s

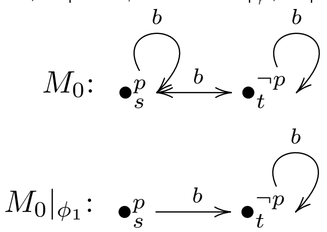
ϕ_0 is not a true lie in M_0, t

ϕ_0 is not a believable true lie in M_0, s

Example: from the outside

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

$$\phi_1 = p \rightarrow B_b(\neg p \rightarrow B_b \neg p)$$
$$= p \rightarrow \neg\hat{B}_b(\neg p \wedge \hat{B}_b p)$$



ϕ_1 is a believable true lie in M_0, s

Example: proper true lie

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

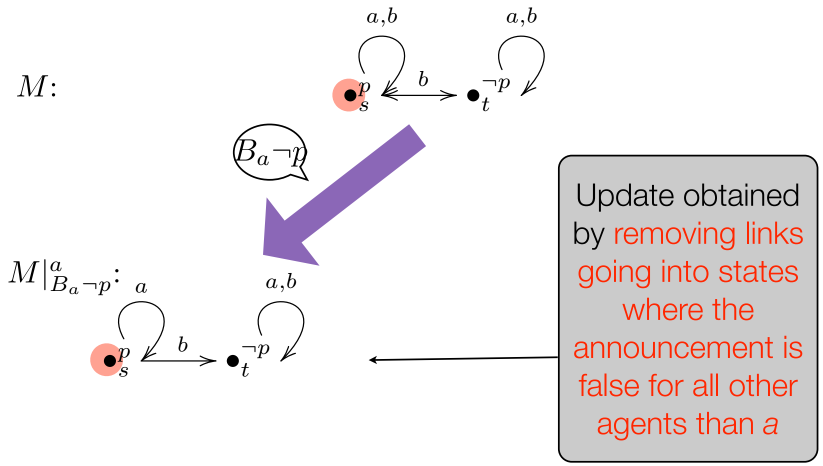
$$\phi_1 = p \rightarrow B_b(\neg p \rightarrow B_b \neg p)$$

Proposition. ϕ_1 *is a true lie in*

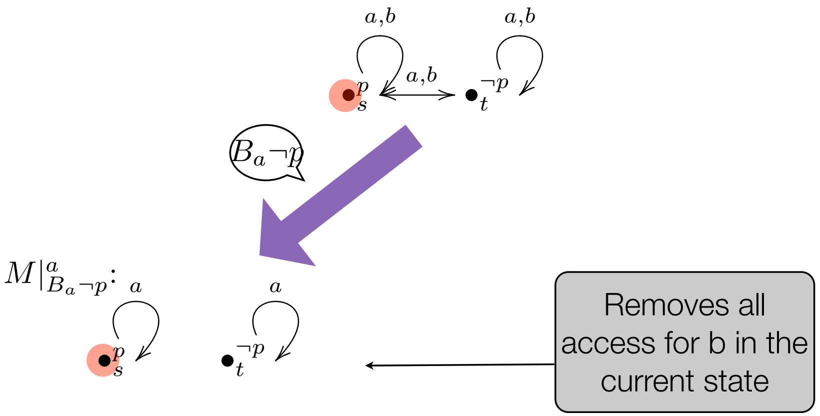
- *KB (the class of all symmetric models)*
- *K45 (the class of all transitive and Euclidian models)*

True lies from the inside

Untruthful announcements by an agent *a* inside the system



Unbelievable lie



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

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

Lie by agent *a*, possible pre-conditions

$\neg \phi$

$\neg B_a \phi$

$B_a \neg \phi$

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

Formalising lies: made by an agent *a* in the system

Given: pointed model M, s

Pre-condition: $M, s \models B_a \neg \phi$

Additional pre-condition for believable lies:
$M, s \models \bigwedge_{i \in Ag} \hat{B}_i \phi$

Consequence: $M|^a_{B_a \phi}, s$ obtained by cutting links to $\neg B_a \phi$-states for all agents $b \neq a$

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

ϕ

$B_a \phi$

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

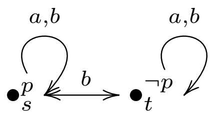
In some model class (typically K(D)45)

Example: from the inside

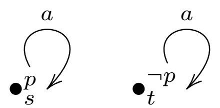
ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$\phi_0 = p \wedge B_b p$

M_0:



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



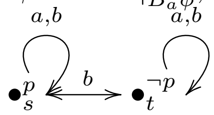
ϕ_0 is a true lie by a in M_0, s
ϕ_0 is not a true lie by a in M_0, t
ϕ_0 is not a believable true lie by a in M_0, s
(it can be shown that ϕ_0 is not a believable true lie on any S5 model)

Example: from the inside

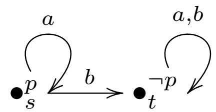
ϕ is a true lie by a in M, s iff $M, s \models B_a \neg\phi$ and $M|^a_{B_a\phi}, s \models \phi$

$$\phi_1 = p \rightarrow B_b(\neg p \rightarrow B_b \neg p)$$
$$= p \rightarrow \neg\hat{B}_b(\neg p \wedge \hat{B}_b p)$$

M_0:



$M_0|^a_{B_a\phi_1}$:



ϕ_1 is a believable true lie by a in M_0, s

Example: proper true lie by *a*

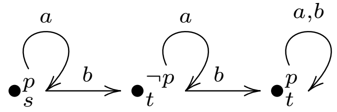
ϕ is a true lie by a iff $\forall M \forall s : \; M, s \models B_a \neg\phi \Rightarrow \quad M|^a_{B_a\phi}, s \models \phi$

$$\phi_1 = p \rightarrow B_b(\neg p \rightarrow B_b \neg p)$$

Proposition. *ϕ_1 is a true lie by any $a \neq b$ in KTB (the class of all reflexive and symmetric models).*

.. but not in K(D)45

$$\phi_1 = p \to B_b(\neg p \to B_b \neg p)$$



Other Moorean phenomena

Relations to (un)successful updates

True lie in M, s:	$M, s \models \neg\phi$ and $M|_\phi, s \models \phi$
Successful update in M, s:	$M, s \models \phi$ and $M|_\phi, s \models \phi$
Unsuccessful update in M, s:	$M, s \models \phi$ and $M|_\phi, s \models \neg\phi$

Other definitions

Self-refuting truth:	$\forall M, s$	$M, s \models \phi$	$\Rightarrow$	$M|_\phi, s \models \neg\phi$
True lie:	$\forall M, s$	$M, s \models \neg\phi$	$\Rightarrow$	$M|_\phi, s \models \phi$
Successful formula:	$\forall M, s$	$M, s \models \phi$	$\Rightarrow$	$M|_\phi, s \models \phi$
Impossible lie:	$\forall M, s$	$M, s \models \neg\phi$	$\Rightarrow$	$M|_\phi, s \models \neg\phi$

Moore sentences again

$$\phi = p \wedge \neg K_b p$$

- Unsuccessful
- Self-refuting

33

Characterisations

- Positive formulae are successful (van Benthem, Visser)

$$\phi ::= p \mid \neg p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \wedge \phi \mid B_i \phi$$

- Complete syntactic characterisation of successful formulae has been an open problem for a long time
- Breakthrough: Holliday and Icard (AiML 2010)
  - Characterises the class of (un)successful as well as self-refuting formulae for the case of one agent only
  - Basic result: “Moorean” phenomena is the source of all unsuccessfulness and self-refutation

34

Open problems

- Holliday and Icard’s result do not carry over to the multi-agents setting, or to agents without negative introspection
  - Non-Moorean unsuccessful formulae exist
- True lies: even more difficult?

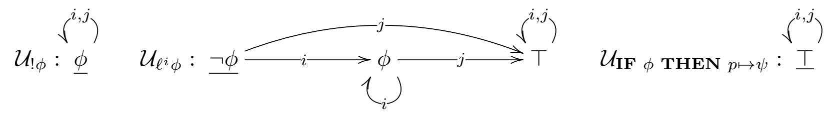
35

On the logic of private true lies

36

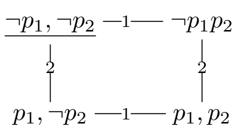
Action models for private lies

$$\phi \ ::= \ \top \mid p \mid \neg\phi \mid \phi \wedge \phi \mid B_i\phi \mid \langle !\phi\rangle\phi \mid \langle \ell^i\phi\rangle\phi \mid \langle \textbf{IF } \phi \textbf{ THEN } p \mapsto \phi\rangle\phi$$

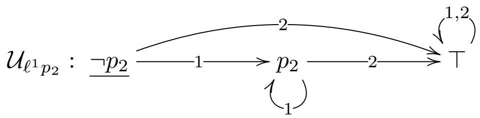


$$
\begin{array}{rcl}
\mathcal{M}, w \vDash p & \Leftrightarrow & p \in V(w) \\
\mathcal{M}, w \vDash \neg\phi & \Leftrightarrow & \mathcal{M}, w \nvDash \phi \\
\mathcal{M}, w \vDash \phi \wedge \psi & \Leftrightarrow & \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \vDash B_i\psi & \Leftrightarrow & \text{for all } v \text{ such that } w \rightarrow_i v : \mathcal{M}_v \vDash \psi \\
\mathcal{M}, w \vDash \langle\star\rangle\phi & \Leftrightarrow & \mathcal{M}, w \vDash Pre(\mathcal{U}_\star) \text{ and } \mathcal{M} \otimes \mathcal{U}_\star, (w,u) \vDash \phi
\end{array}
$$

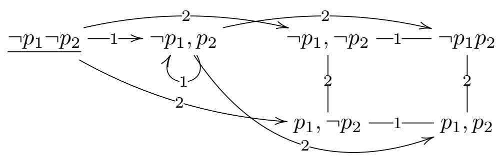
The party example



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

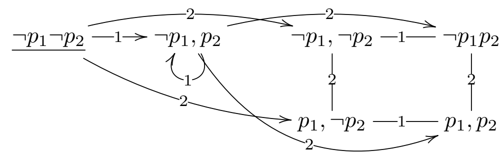


Updated model $(\mathcal{M} \otimes \mathcal{U})$

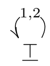


Example (continued)

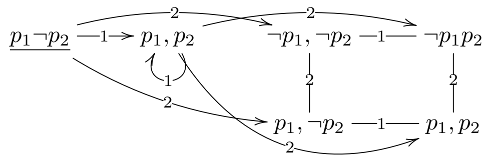
Updated model $(\mathcal{M} \otimes \mathcal{U})$



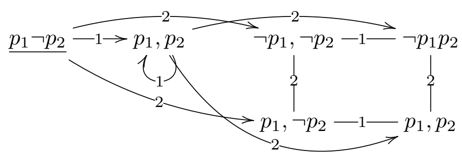
The update model $\mathcal{U}'$ for **IF** $B_1 p_2$ **THEN** $p_1 \mapsto \top$:



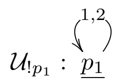
Updated model $(\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}')$



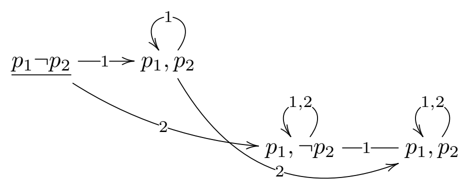
Example (continued)



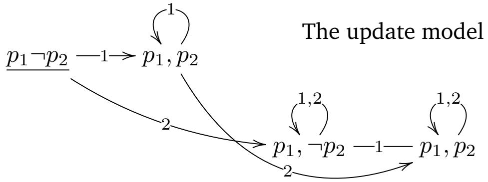
The update model $\mathcal{U}''$ for $!p_1$:



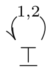
Updated model $(\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}' \otimes \mathcal{U}'')$



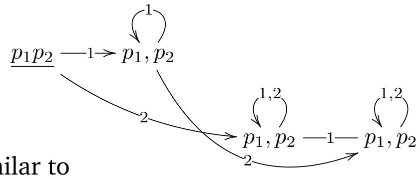
Example (continued)



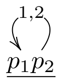
The update model $\mathcal{U}'''$ for **IF** $B_2 p_1$ **THEN** $p_2 \mapsto \top$:



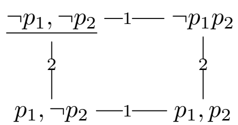
Updated model $(\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}' \otimes \mathcal{U}'' \otimes \mathcal{U}''')$



which is similar to



Example (continued)



$\mathcal{M}, w \models \neg p_1 \wedge \neg p_2 \wedge \langle \ell_1 p_2 \rangle \langle \textbf{IF } B_1 p_2 \textbf{ THEN } p_1 \mapsto \top \rangle \langle ! p_1 \rangle \langle \textbf{IF } B_2 p_1 \textbf{ THEN } p_2 \mapsto \top \rangle p_1 \wedge p_2 \wedge B_{1,2}(p_1 \wedge p_2)$

Summary

- Formalised true lies
- Many subtleties
- Related to other Moorean phenomena
- Characterisation is hard
- Future work:
    - Understanding relationships
    - Lying games