



HOKKAIDO UNIVERSITY

Title	True Lies
Author(s)	Ågotnes, Thomas
Description	SOCREAL 2013 : 3rd International Workshop on Philosophy and Ethics of Social Reality 2013. Hokkaido University, Sapporo, Japan, 25-27 October 2013. Session 1 : Logic, Knowledge, and Philosophy of Language
Citation	Proceedings of SOCREAL 2013 : 3rd International Workshop on Philosophy and Ethics of Social Reality 2013, 1-11
Issue Date	2013-10-25
Doc URL	https://hdl.handle.net/2115/55072
Type	conference paper
File Information	01Thomas_truelies-4p_rev.pdf



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



1

Background

3

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

2

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:

$$M = (S, \sim_1, \dots, \sim_n, V) \quad \sim_i \text{ equivalence rel. over } S$$

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

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

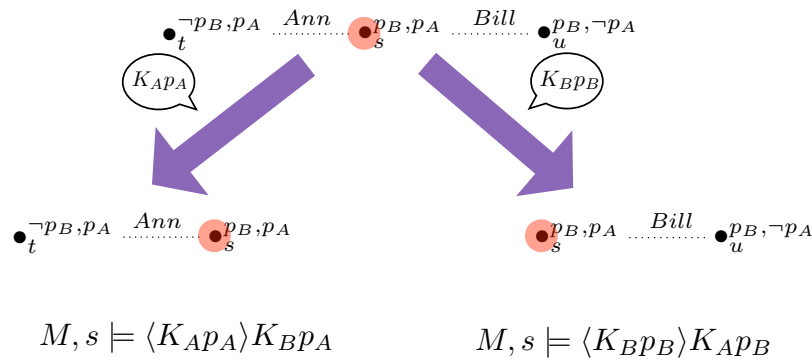
The model resulting from removing states where ϕ_1 is false

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

1

4

Example

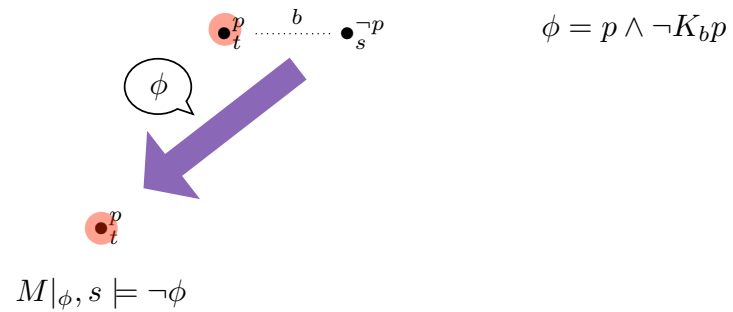


5

Formalising true lies

7

Moore sentences



6

Lies

• Dimensions:

- Who is the liar: 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
 - ...

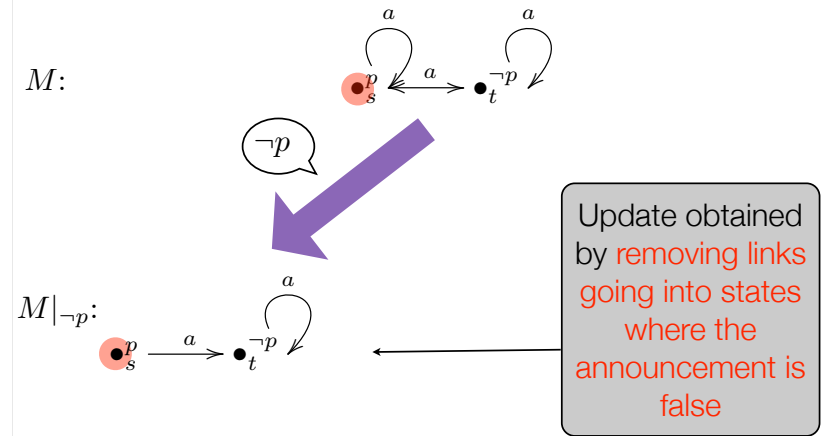
8

Lies

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

9

Untruthful announcements: link-cutting semantics

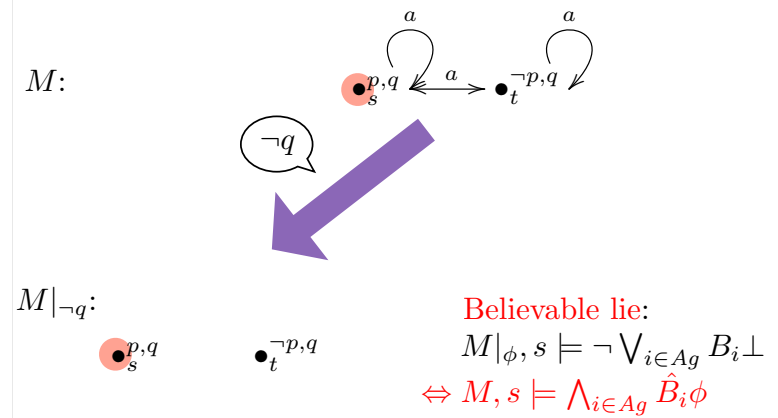


11

True lies from the outside

10

Unbelievable lie



12

Models of lying

Already seen:

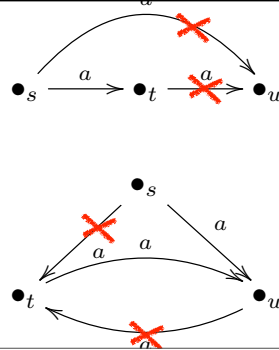
- reflexivity is not preserved under lying
- seriality preserved only for

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

Preservation of transitivity:

Preservation of seriality:

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



13

True lies: from the outside

ϕ is a **believable** true lie in M, s iff $M, s \models \neg\phi$ and $M|_{\phi}, s \models \phi$ and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b \phi$

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

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

15

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

14

Example: from the outside

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



$M_0|_{\phi_0}$: $\bullet_s^p$ $\bullet_t^{\neg p}$

ϕ_0 is a true lie in M_0, s

ϕ_0 is not a true lie in M_0, t

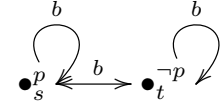
ϕ_0 is not a **believable** true lie in M_0, s

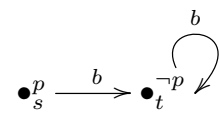
16

Example: from the outside

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

$$\begin{aligned}\phi_1 &= p \rightarrow B_b(\neg p \rightarrow B_b\neg p) \\ &= p \rightarrow \neg\hat{B}_b(\neg p \wedge \hat{B}_bp)\end{aligned}$$

M_0 : 

$M_0|_{\phi_1}$: 

ϕ_1 is a believable true lie in M_0, s

17

True lies from the inside

19

Example: proper true lie

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

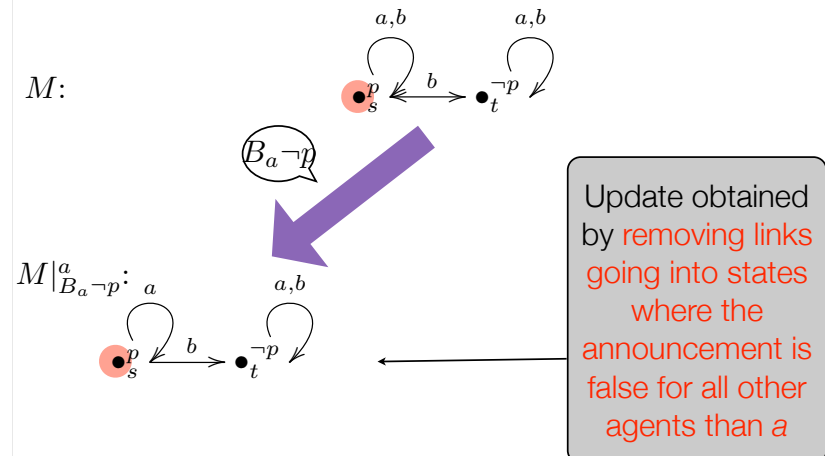
$$\phi_1 = p \rightarrow B_b(\neg p \rightarrow B_b\neg p)$$

Proposition. ϕ_1 is a true lie in

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

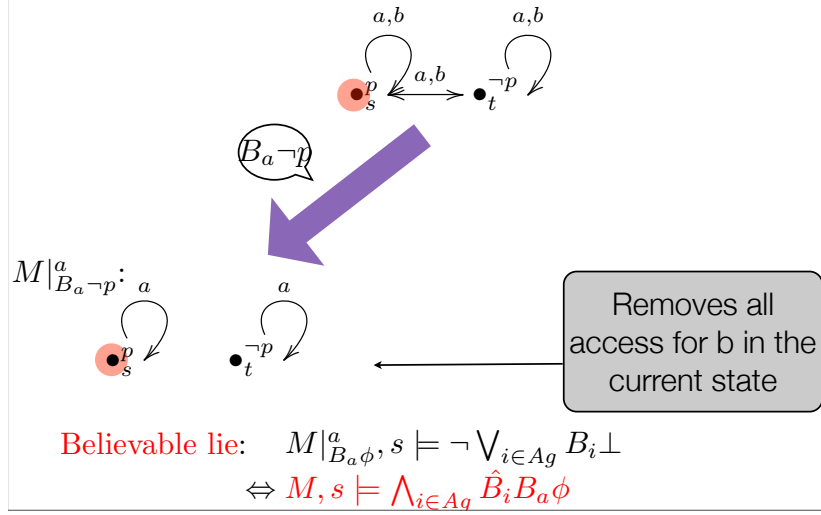
18

Untruthful announcements by an agent a inside the system



20

Unbelievable lie



21

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

23

Lie by agent a , possible pre-conditions

$\neg \phi$

$\neg B_a \phi$

$B_a \neg \phi$

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

22

True lie by agent a , possible post-conditions

ϕ

$B_a \phi$

24

True lies: from the inside

ϕ is a **true lie by a in M, s** iff $M, s \models B_a \neg \phi$ and $M|_{B_a \phi}^a, s \models \phi$
 and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi$

ϕ is a **true lie by a** iff $\forall M \forall s : (M, s \models B_a \neg \phi \Rightarrow M|_{B_a \phi}^a, s \models \phi$
 and $M, s \models \bigwedge_{b \in Ag} \hat{B}_b B_a \phi)$

believable

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

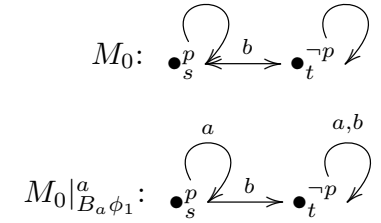
25

Example: from the inside

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

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

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

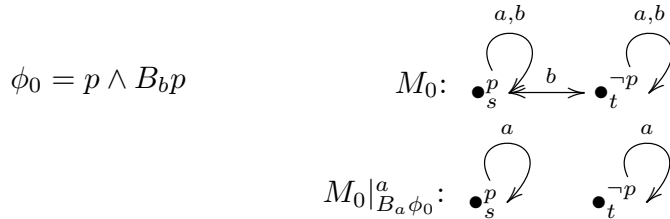


ϕ_1 is a believable true lie by a in M_0, s

27

Example: from the inside

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



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)

26

Example: proper true lie by a

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

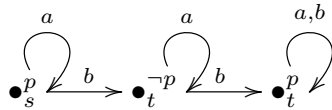
$$\phi_1 = p \rightarrow B_b(\neg p \rightarrow B_b \neg p)$$

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

28

.. but not in K(D)45

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



29

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

31

Other Moorean phenomena

30

Other definitions

Self-refuting truth: $\forall M, s \quad M, s \models \phi \Rightarrow M|_\phi, s \models \neg\phi$

True lie: $\forall M, s \quad M, s \models \neg\phi \Rightarrow M|_\phi, s \models \phi$

Successful formula: $\forall M, s \quad M, s \models \phi \Rightarrow M|_\phi, s \models \phi$

Impossible lie: $\forall M, s \quad M, s \models \neg\phi \Rightarrow M|_\phi, s \models \neg\phi$

32

Moore sentences again

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

- Unsuccessful
- Self-refuting

33

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

Characterisations

- Positive formulae are successful (van Benthem, Visser)
$$\phi ::= p \mid \neg p \mid \neg \phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid B_i \phi$$
- 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

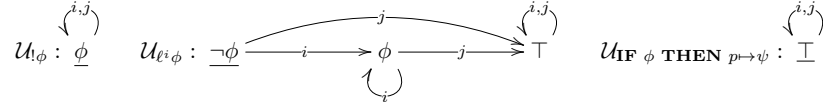
34

On the logic of private true lies

36

Action models for private lies

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



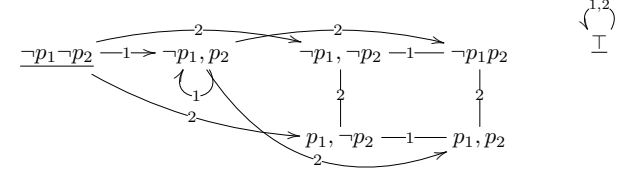
$\mathcal{M}, w \models p$	$\Leftrightarrow$	$p \in V(w)$
$\mathcal{M}, w \models \neg\phi$	$\Leftrightarrow$	$\mathcal{M}, w \not\models \phi$
$\mathcal{M}, w \models \phi \wedge \psi$	$\Leftrightarrow$	$\mathcal{M}, w \models \phi$ and $\mathcal{M}, w \models \psi$
$\mathcal{M}, w \models B_i\psi$	$\Leftrightarrow$	for all v such that $w \rightarrow_i v : \mathcal{M}_v \models \psi$
$\mathcal{M}, w \models \langle \star \rangle\phi$	$\Leftrightarrow$	$\mathcal{M}, w \models \text{Pre}(\mathcal{U}_\star)$ and $\mathcal{M} \otimes \mathcal{U}_\star, (w, u) \models \phi$

37

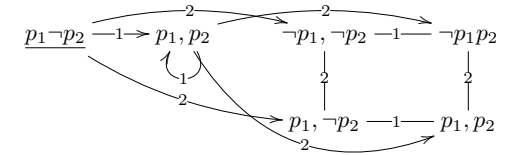
Example (continued)

The update model $\mathcal{U}'$ for **IF** B_1p_2 **THEN** $p_1 \mapsto \top$:

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



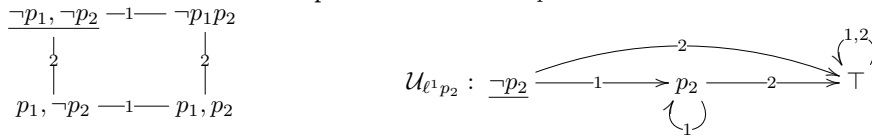
Updated model $(\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}')$



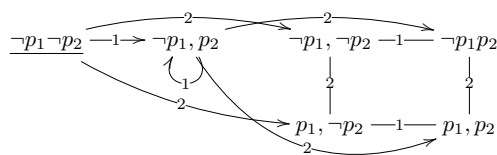
39

The party example

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



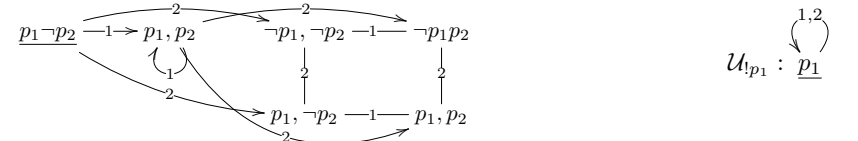
Updated model $(\mathcal{M} \otimes \mathcal{U})$



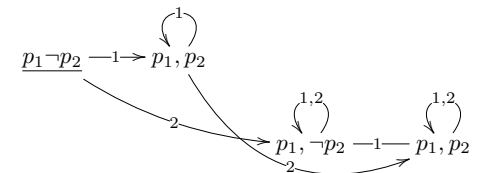
38

Example (continued)

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

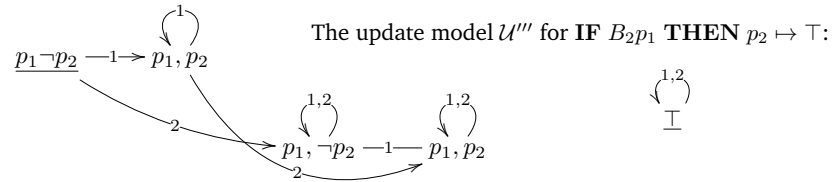


Updated model $(\mathcal{M} \otimes \mathcal{U} \otimes \mathcal{U}' \otimes \mathcal{U}'')$

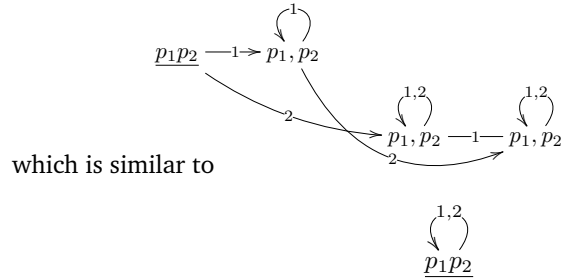


40

Example (continued)



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



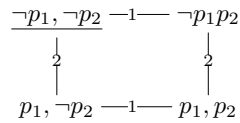
41

Summary

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

43

Example (continued)



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

42