

Lies, deception, and computation

Hans van Ditmarsch, LORIA, CNRS, Université de Lorraine, France

Presentation for frAlday, 6 November 2020

IN A WORLD WHERE EVERYONE CAN ONLY
TELL THE TRUTH,



...HE'S JUST
INVENTED
THE LIE!

Ricky Gervais Jennifer Garner Jonah Hill Louis C.K. with Rob Lowe and Tina Fey

The **Invention of Lying**

THE INVENTION OF LYING: A FILM BY LIAM NEWMAN. CASTING BY JENNIFER COOPER. COSTUME DESIGNER: JENNIFER COOPER. HAIR AND MAKEUP: JENNIFER COOPER. PRODUCTION DESIGNER: JENNIFER COOPER. EXECUTIVE PRODUCERS: JENNIFER COOPER, JENNIFER COOPER, JENNIFER COOPER. PRODUCED BY JENNIFER COOPER. WRITTEN BY JENNIFER COOPER. DIRECTED BY LIAM NEWMAN. CASTING BY JENNIFER COOPER. COSTUME DESIGNER: JENNIFER COOPER. HAIR AND MAKEUP: JENNIFER COOPER. PRODUCTION DESIGNER: JENNIFER COOPER. EXECUTIVE PRODUCERS: JENNIFER COOPER, JENNIFER COOPER, JENNIFER COOPER. PRODUCED BY JENNIFER COOPER. WRITTEN BY JENNIFER COOPER. DIRECTED BY LIAM NEWMAN.

www.inventionoflying.com IN CINEMAS 2 OCTOBER 2009

Lying in a dynamic epistemic logic

What is a lie?

- ▶ You are lying if you say that φ but you believe that $\neg\varphi$.
- ▶ For me to believe the lie that φ , I must consider φ possible.

What is the truth?

- ▶ You are truthful if you say that φ and you believe that φ .
- ▶ For me to believe the truth that φ , I must consider φ possible.

Public announcement logic

- ▶ A lie is an epistemic action transforming Kripke models.
- ▶ Modelling lies is not so different from modelling the truth.
- ▶ *Public announcement logic* [Plaza 1989] models the truth.

Consecutive numbers

Anne and Bill are each going to be told a natural number. Their numbers will be one apart. The numbers are now being whispered in their respective ears. They are aware of this scenario. Suppose Anne is told 2 and Bill is told 3.

The following truthful conversation between Anne and Bill now takes place:

- ▶ *Anne: "I do not know your number."*
- ▶ *Bill: "I do not know your number."*
- ▶ *Anne: "I know your number."*
- ▶ *Bill: "I know your number."*

Explain why is this possible.

Consecutive numbers — representing uncertainties

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) - \dots$

$(0,1) - b - (2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) - \dots$

$(0,1) - b - (2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

- ▶ Anne: “I do not know your number.” ??

Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) - \dots$

$(0,1) - b - (2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

- ▶ Anne: “I do not know your number.” **eliminated states**

Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) - \dots$

$(2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

- ▶ Anne: "I do not know your number."

Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) - \dots$

$(2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.” ??

Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) - \dots$

$(2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.” **eliminated states**

Consecutive numbers — successive announcements

$$(1,2) - b - (3,2) - a - (3,4) - \dots$$
$$\underline{(2,3)} - b - (4,3) - \dots$$

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.”

Consecutive numbers — successive announcements

$$(1,2) - b - (3,2) - a - (3,4) - \dots$$
$$\underline{(2,3)} - b - (4,3) - \dots$$

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.”
- ▶ Anne: “I know your number.” ??

Consecutive numbers — successive announcements

$(1,2) - b - (3,2) - a - (3,4) - \dots$

$(\underline{2},3) - b - (4,3) - \dots$

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.”
- ▶ Anne: “I know your number.” **eliminated states**

Consecutive numbers — successive announcements

(1,2)

(2,3)

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.”
- ▶ Anne: “I know your number.”

Consecutive numbers — successive announcements

(1,2)

(2,3)

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.”
- ▶ Anne: “I know your number.”
- ▶ Bill: “I know your number.” ??

Consecutive numbers — successive announcements

(1,2)

(2,3)

- ▶ Anne: “I do not know your number.”
- ▶ Bill: “I do not know your number.”
- ▶ Anne: “I know your number.”
- ▶ Bill: “I know your number.” **already common knowledge**

Consecutive numbers with lying — initial model, once more

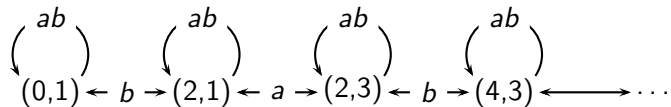
$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) - \dots$

$(0,1) - b - (2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

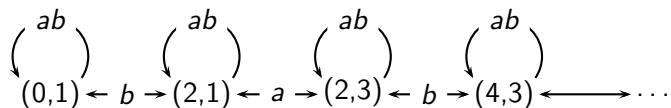
Consecutive numbers with lying — connected submodel

$(0,1) - b - (2,1) - a - \underline{(2,3)} - b - (4,3) - \dots$

Consecutive numbers with lying — same with all arrows

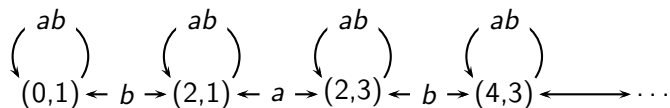


Consecutive numbers with lying — same with all arrows

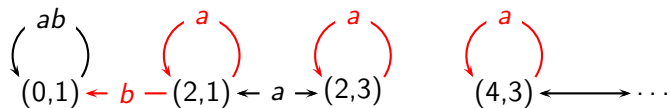


▶ Anne: "I know your number." Anne is lying

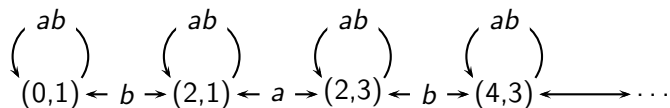
Consecutive numbers with lying — same with all arrows



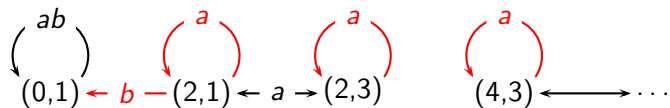
► Anne: "I know your number." Anne is lying



Consecutive numbers with lying — same with all arrows



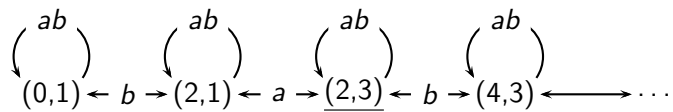
► Anne: "I know your number." Anne is lying



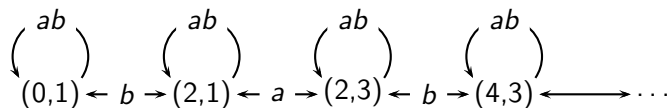
► Bill: "That's a lie."



Consecutive numbers with lying

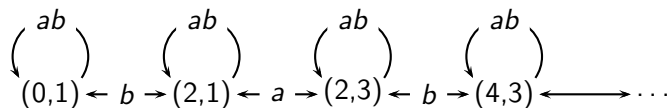


Consecutive numbers with lying

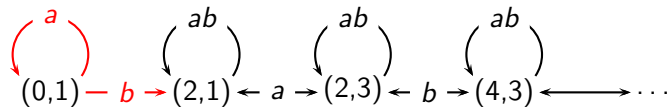


▶ Anne: "I do not know your number."

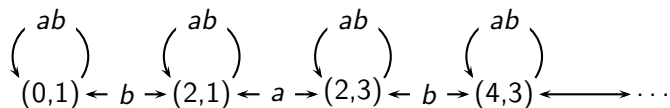
Consecutive numbers with lying



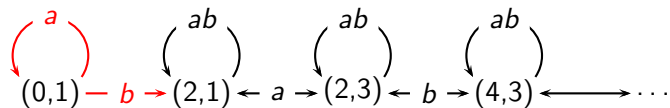
► Anne: "I do not know your number."



Consecutive numbers with lying

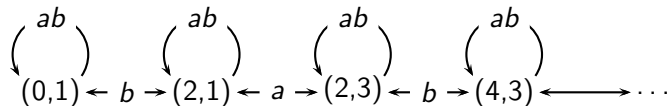


► Anne: "I do not know your number."

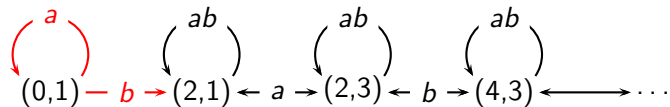


► Bill: "I know your number." **Bill is lying**

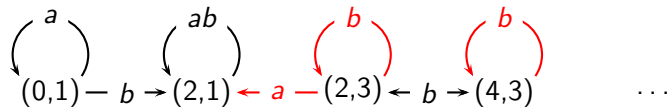
Consecutive numbers with lying



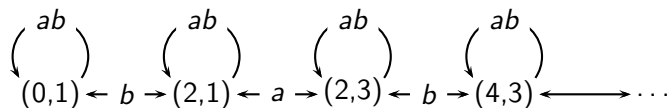
► Anne: "I do not know your number."



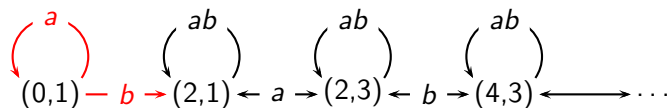
► Bill: "I know your number." **Bill is lying**



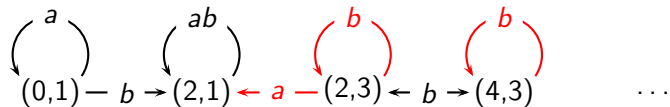
Consecutive numbers with lying



► Anne: "I do not know your number."

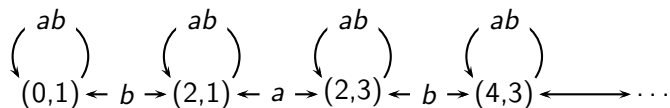


► Bill: "I know your number." **Bill is lying**

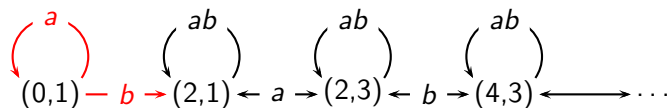


► Anne: "I know your number." **Anne is mistaken.**

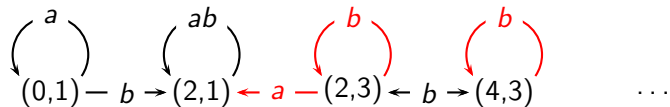
Consecutive numbers with lying



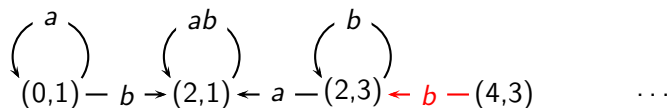
► Anne: "I do not know your number."



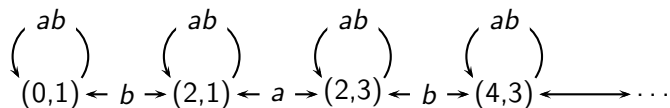
► Bill: "I know your number." **Bill is lying**



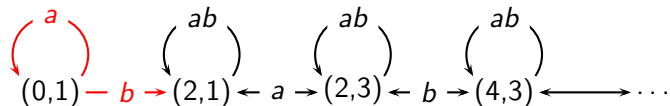
► Anne: "I know your number." **Anne is mistaken.**



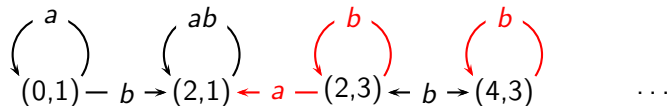
Consecutive numbers with lying



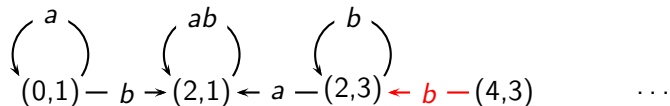
► Anne: "I do not know your number."



► Bill: "I know your number." **Bill is lying**



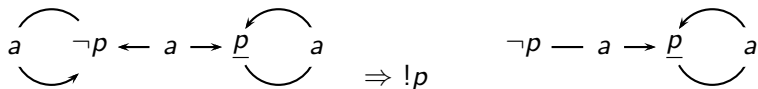
► Anne: "I know your number." **Anne is mistaken.**



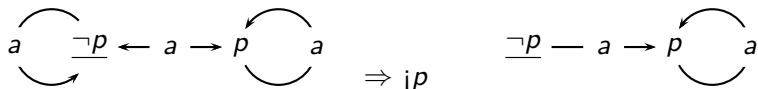
► Bill: "I know your number." **True, but not what Anne thinks.**
There are no informative consequences of this announcement.

Semantics of truth and lies in dynamic epistemic logic

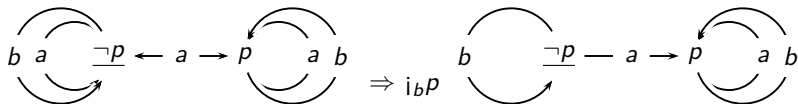
- ▶ The truth that φ restricts a model to states where φ is true.
... restricts it to arrows (pointing to states) where φ is true.



- ▶ The lie that φ **also** restricts arrows to states where φ is true.



- ▶ The lie by b that φ restricts arrows for other agents a to states where φ is true (or where φ is believed by b to be true).



- ▶ Bluffing, skeptical agents, belief revising agents, deception, ...

Oranges freeze in Sevilla = p ; Hans = a ; you = b



- ▶ Truthful announcement that p : $B_a p$ and $!_a p$
- ▶ Lying announcement that p : $B_a \neg p$ and $!_a \neg p$
- ▶ Bluffing announcement that p : $\neg(B_a p \vee B_a \neg p)$ and $!_a p$
- ▶ Honest mistake that p : $\neg p \wedge B_a p$ en $!_a p$

References

- ▶ HvD, Jan van Eijck, Yanjing Wang, Floor Sietsma. *On the logic of lying*. LNCS 7010, pp. 41–72, 2012.
- ▶ HvD, *Dynamics of Lying*, Synthese 191(5): 745–777, 2014.
- ▶ Thomas Ågotnes, HvD, Yanjing Wang. *True lies*. Synthese 195(10): 4581–4615, 2018.
- ▶ HvD, Petra Hendriks, Rineke Verbrugge. *Editors' Review and Introduction: Lying in Logic, Language, and Cognition*. topiCS 12(2): 466–484, 2020.

Private Lies

True Lies and Butterflies

Mei wants to invite two friends (Zhu) Yingtai and (Liang) Shanbo to a party. She knows that they are dying to get close to each other. Thus one will come if and only if (s)he believes that the other will come. Obviously they do not yet wish to admit this to each other, because as far as they know they are both still very uncertain about each other's feelings. Given this uncertainty, both in fact don't intend to come to the party. **Mei now lies to Yingtai that Shanbo will come to the party and she lies to Shanbo that Yingtai will come to the party.** As a result, they will both come to the party.

(Adaptation of the *Butterfly Lovers*, a Chinese folktale, see https://en.wikipedia.org/wiki/Butterfly_Lovers)

The Butterfly Lovers



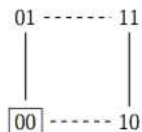
Private Lies

Mei: lies to Yingtai and to Shanbo

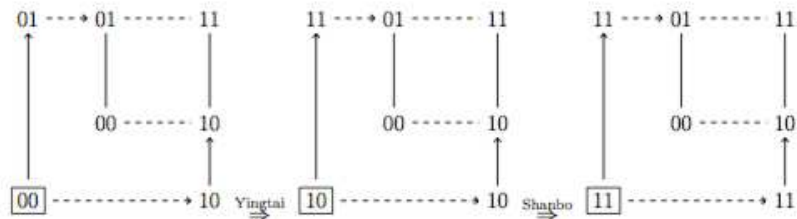
Yingtai: changes her mind

Shanbo: changes his mind

(01: Yingtai does not go and Shanbo goes, etc.)



⇓ Mei



Computational complexity of lying and deception

Modal logical analysis of lying result in more complex models. With degrees of belief, or belief and knowledge combined, models are even more complex. Both for the liar and for the listener.

- ▶ The liar has to remember the lie and the truth, and remain consistent
- ▶ The skeptical listener has to remember the supposed truth and the suspected truth

This has been demonstrated in many psychological experiments. Is there a parallel in computational complexity?

Possible computational benchmarks

- ▶ Complexity of **model checking** with lying and deception
- ▶ Complexity of **satisfiability** with lying and deception
- ▶ Complexity of **consistency** checking in evolving communication
- ▶ Complexity of **epistemic planning** with lying and deception
- ▶ Deception by omission in **asynchronous** distributed computing

Complexity of lying and deception — game theory

Complex modal languages and structures seem lack realism. Often lying *simplifies* decision making. It is a shortcut, lowering computational / brain requirements. The liar says what the listener wants to hear; such lying in job interviews:

Glazer & Rubinstein, *A Model of Persuasion with Boundedly Rational Agents*, Journal of Political Economy, 2012.

Cheating is most important in *iterated games*. The cheat fears the *negative payoff*, but mostly *no longer being allowed to play*: **trust**.

Possible benchmarks

- ▶ **epistemic** *Ulam games*: truth tracking in the presence of a single lie or a single liar; robustness
- ▶ infinite games with lying; **distinguishing lies from mistakes**
- ▶ gossip protocols with lying; **distinguishing lies from noise** (Malvin Gattinger, Line van den Berg)
- ▶ (higher-order) theory of mind modellings of deception (Rineke Verbrugge), omission, misdirection (Frank Dignum)