Dynamic epistemic logic and lying

Hans van Ditmarsch
LORIA — CNRS / Univ. de Lorraine & (associate) IMSc, Chennai
hans.van-ditmarsch@loria.fr
http://personal.us.es/hvd/


ANTON PIECK

The Ditmarsch Tale of Wonders

I will tell you something. I saw two roasted fowls flying; they flew quickly and had their breasts turned to Heaven and their backs to Hell; and an anvil and a mill-stone swam across the Rhine prettily, slowly, and gently; and a frog sat on the ice at Whitsuntide and ate a ploughshare.

. . .

Open the window that the lies may fly out.

Jacob Ludwig Grimm and Wilhelm Carl Grimm, Fairy Tales

'The Ditmarsch Tale of Wonders' is called in German:
'Das Dietmarsische Lügenmärchen'.

Lying and truth telling

In the Grimm Brothers fairy tale it is clear that the speaker lies.
Can you lie without the listener noticing?
What are the informative consequences of lying?

What are the informative consequences of telling the truth?
Let us recall **public announcement logic**.

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.

Oldest known source:
Littlewood, A Mathematician's Miscellany, 1953

Consecutive numbers — representing uncertainties

$$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \text{———} \cdots$$

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

Consecutive numbers — successive announcements

$(1{,}0) - a - (1{,}2) - b - (3{,}2) - a - (3{,}4) \text{———} \cdots$

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

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

Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \text{———} \cdots$

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

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

Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \text{———} \cdots$

$(2,1) - a - \underline{(2,3)} - b - (4,3) \text{———} \cdots$

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

Consecutive numbers — successive announcements

$$(1{,}0) - a - (1{,}2) - b - (3{,}2) - a - (3{,}4) \text{———} \cdots$$

$$(2{,}1) - a - \underline{(2{,}3)} - b - (4{,}3) \text{———} \cdots$$

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

Consecutive numbers — successive announcements

$$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \;\text{———}\; \cdots$$

$$(2,1) - a - \underline{(2,3)} - b - (4,3) \;\text{———}\; \cdots$$

- 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) ———— …

<u>(2,3)</u> — b — (4,3) ———— …

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

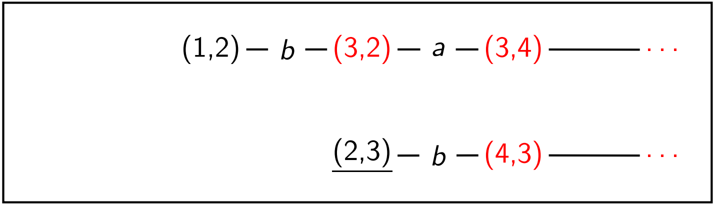
Consecutive numbers — successive announcements

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

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

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

Consecutive numbers — successive announcements



- 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)

<u>(2,3)</u>

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

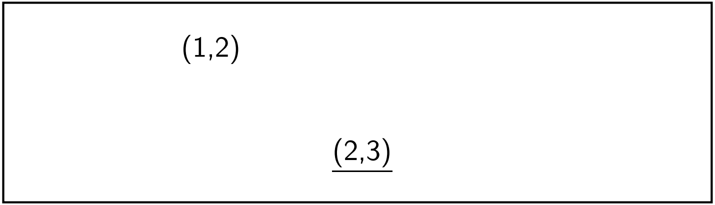
Consecutive numbers — successive announcements

(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



- 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 — 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."

Truthful (true) public announcement logic

Jan Plaza, Logics of public communications, 1989 & 2007

- **Structures**: pointed Kripke models
  E.g., $\underline{(0,1)}$—b—$(2,1)$—a—$(2,3)$—...
- **Language**: $p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_a\varphi \mid [!\varphi]\varphi$
  $B_a\varphi$ for 'agent a believes/knows φ';
  $[!\varphi]\psi$ for 'after truthful announcement of φ, ψ is true'.
- **Semantics**:
  $B_a\varphi$ is true in a state iff: φ is true in all a-accessible states.
  In state $(0,1)$ formula $B_b 1_b$ (b knows his number is 1') is true.
  $[!\varphi]\psi$ is true in a state iff: whenever φ is true, in the restriction of the model to the φ-states, ψ is true.
  In state $(0,1)$ formula $\neg B_b 0_a \wedge [!0_a] B_b 0_a$ is true.

Lying in public announcement logics

Let us start with an example...

Consecutive numbers, with lying

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

- 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, with lying

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

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

- Anne: “I know your number.” Anne is lying

Consecutive numbers, with lying

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

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

- Anne: "I know your number." Anne is lying
- Bill: "You're lying."

Consecutive numbers, with lying

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

- 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, with lying

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

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

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

Consecutive numbers, with lying

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

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

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

Consecutive numbers, with lying

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

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

- Anne: "I do not know your number."
- Bill: "I know your number." Bill is lying

Consecutive numbers, with lying

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

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

- Anne: "I do not know your number."
- Bill: "I know your number." Bill is lying
- Anne: "I know your number."

Consecutive numbers, with lying

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

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

- Anne: "I do not know your number."
- Bill: "I know your number." Bill is lying
- Anne: "I know your number." Anne is mistaken.
  Anne *thinks* to know that Bill has 1.

Consecutive numbers, with lying

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

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

- Anne: “I do not know your number.”
- Bill: “I know your number.” Bill is lying
- Anne: “I know your number.” Anne is mistaken. Anne *thinks* to know that Bill has 1.
- (Bill: “I know your number.” By now, this is true!)

What is a lie?

- You are lying if you say to me that φ (is true), but believe that $\neg\varphi$ (is true). (With the *intention* for me to believe φ.)
- The lie was effective (the intention has been successfully realized) if I now believe that φ was true. ('Was', not 'is', for technical reasons.)
- For me to believe your lie that φ, I must consider it possible that φ. (Otherwise I will believe that you're lying!)

Lying by an outside observer *Lying public announcement*

- The agents are the listeners whose beliefs are modelled.
- Lies are announcements made by an outsider (not modelled).
- The announcements are always believed.

Example of truthful public announcement

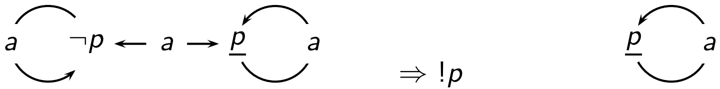
Let p be the proposition ‘Oranges freeze in Sevilla’.
Agent a does not know whether this is true.
This uncertainty can be modelled as follows:

$$\neg p \text{ — } a \text{ — } \underline{p}$$

After the announcement of p we get:

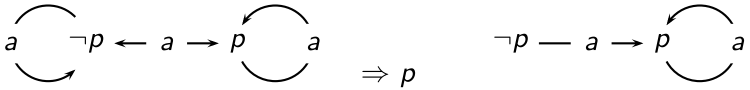
$$\neg p \text{ — } a \text{ — } \underline{p} \qquad \Rightarrow !p \qquad \underline{p}$$

To model lying, later on, we need a more explicit visualization:



A different semantics: Believed announcements

An alternative to the logic of truthful public announcements is the logic of *believed (public) announcements.* The effect of the announcement of φ, is that only states where φ is true remain accessible for the agents. The announcement may be false.



After the announcement, *a* believes that oranges freeze in Sevilla. No matter what the truth is.

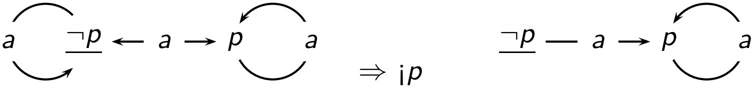
Believed announcements are investigated in:

- Jelle Gerbrandy, Bisimulations on Planet Kripke, ILLC 1999
- Barteld Kooi, Expressivity (...) via reduction axioms. Journal of Applied Non-Classical Logics 17(2): 231-253, 2007

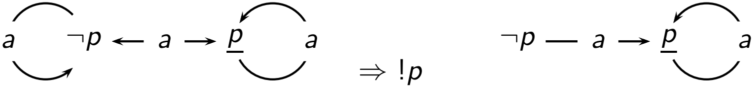
Lie as Lying public announcement

In case of a lie that φ: φ is false; it is announced that φ is true; after the announcement, agent a believes that φ (was true).

Lying public announcement for which we write $¡\varphi$ is one execution of believed announcement:



Truthful public announcement $!\varphi$ is another execution of believed announcement:



Principles of lying public announcement

Axioms for truthful public announcement:

$$\begin{array}{rcl} [!\varphi]p & \leftrightarrow & \varphi \rightarrow p \\ [!\varphi]\neg\psi & \leftrightarrow & \varphi \rightarrow \neg[!\varphi]\psi \\ [!\varphi](\psi_1 \wedge \psi_2) & \leftrightarrow & [!\varphi]\psi_1 \wedge [!\varphi]\psi_2 \\ [!\varphi]B_i\psi & \leftrightarrow & \varphi \rightarrow B_i[!\varphi]\psi \end{array}$$

Dual axioms for lying:

$$\begin{array}{rcl} [¡\varphi]p & \leftrightarrow & \neg\varphi \rightarrow p \\ [¡\varphi]\neg\psi & \leftrightarrow & \neg\varphi \rightarrow \neg[¡\varphi]\psi \\ [¡\varphi](\psi_1 \wedge \psi_2) & \leftrightarrow & [¡\varphi]\psi_1 \wedge [¡\varphi]\psi_2 \\ [¡\varphi]B_i\psi & \leftrightarrow & \neg\varphi \rightarrow B_i[!\varphi]\psi \end{array}$$

Combined, the principles deliver: (where $[\varphi]\psi \leftrightarrow ([!\varphi]\psi \wedge [¡\varphi]\psi)$)

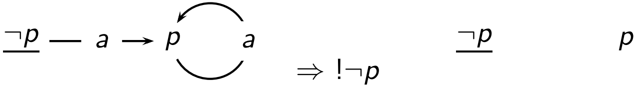
$$[\varphi]B_i\psi \quad \leftrightarrow \quad B_i(\varphi \rightarrow [\varphi]\psi)$$

Principles of lying public announcement

$$[¡\varphi]B_i\psi \quad \leftrightarrow \quad \neg\varphi \rightarrow B_i[!\varphi]\psi$$

(After the lie that φ, agent i believes that ψ,) iff, (on condition that φ is false, agent i believes that ψ after truthful announcement that φ).

In lying (and truthful) public announcement agents may go 'crazy' (empty access / believe inconsistencies). This is a problem.



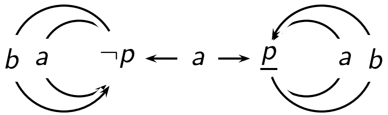
Going crazy can be avoided elegantly by requiring that the listeners consider it possible that the lie is true: preconditions $\neg B_a \neg\varphi$.

- Hans van Ditmarsch, Jan van Eijck, Yanjing Wang, Floor Sietsma. *On the logic of lying*, LNCS 7010, pp. 41-72, 2012.

Lying public announcement to lying agent announcement

In *lying public announcement* it is implicit that the speaker believes that the announcement is false. We can make this explicit. The result is the logic of *(lying) agent announcement*.

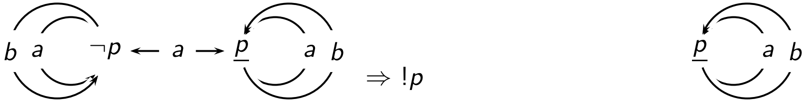
Consider the information state where a does not know whether p, b knows whether p, and p is true.



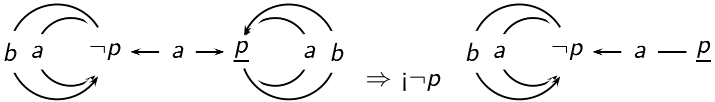
Oranges freeze in Sevilla.
Bill (b) knows whether this is true.
Anne (a) is ignorant.
(And this is common knowledge.)

Lying agent announcement — example

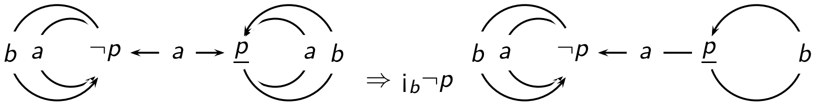
Clearly, a public announcement is not a lie from b to a.



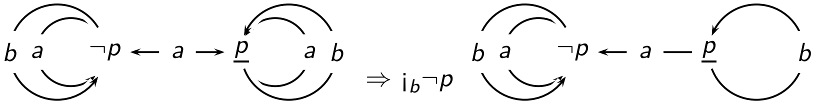
A public lie is also not a lie from b to a.



Instead, a lie from b to a should have the following effect:



Lying agent announcement — example



After agent b lies to a that $\neg p$, we have that:

- b still believes that p;
- a believes that $\neg p$;
- a believes that b believes $\neg p$;
- (a believes that a and b have common belief of $\neg p$.)

Semantics and principles of agent announcement

- The accessibility relation for speaker b does not change.
- The accessibility relation for listener a changes: only the states where (speaker) b believes φ remain accessible for a.

Preconditions of agent announcements (by b) that φ

- Truthful agent announcement $!_b\varphi$: $B_b\varphi$
- Lying agent announcement $¡_b\varphi$: $B_b\neg\varphi$
- Bluffing agent announcement $¡!_b\varphi$: $\neg(B_b\varphi \vee B_b\neg\varphi)$

Principles for b lying to a that φ
(abbreviation $[_b\varphi]\psi \leftrightarrow ([!_b\varphi]\psi \wedge [¡_b\varphi]\psi \wedge [¡!_b\varphi]\psi)$)

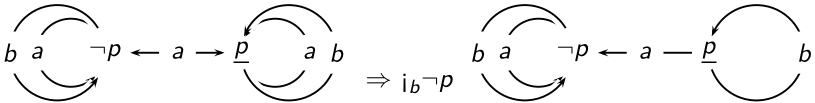
$$[¡_b\varphi]B_a\psi \quad\leftrightarrow\quad B_b\neg\varphi \rightarrow B_a[!_b\varphi]\psi$$
$$[¡_b\varphi]B_b\psi \quad\leftrightarrow\quad B_b\neg\varphi \rightarrow B_b[_b\varphi]\psi$$

$$[¡!_b\varphi]B_a\psi \quad\leftrightarrow\quad \neg(B_b\varphi \vee B_b\neg\varphi) \rightarrow B_a[!_b\varphi]\psi$$

...

When speaker b is caught as a liar

This lie is believed:



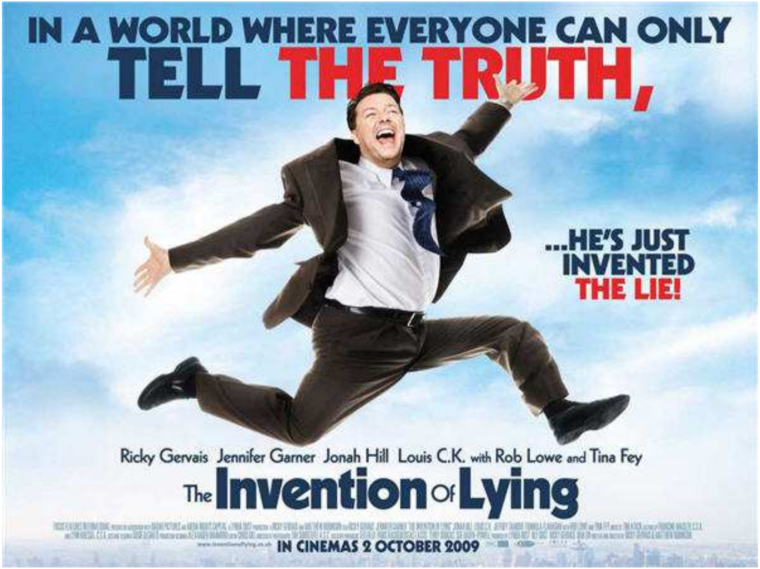
This lie should not be believed:

Agent a now believes ‘everything’. (There are no arrows for a.)

We can elegantly solve this by strengthening the precondition to $\neg B_a \neg B_b \varphi$ (listener a considers possible that speaker b believes φ).

- Hans van Ditmarsch, *Dynamics of lying*, Synthese, 2013

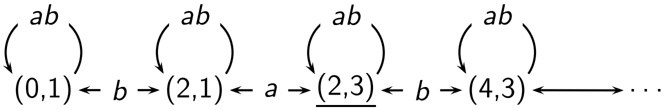
The invention of lying



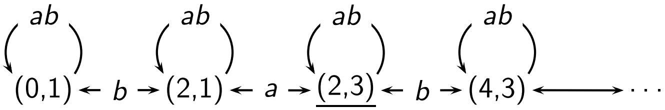
Consecutive numbers with lying

$(0,1) — b — (2,1) — a — \underline{(2,3)} — b — (4,3) \text{———} \cdots$

Consecutive numbers with lying

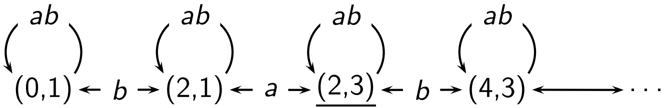


Consecutive numbers with lying

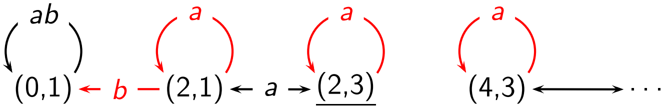


- Anne: "I know your number." Anne is lying

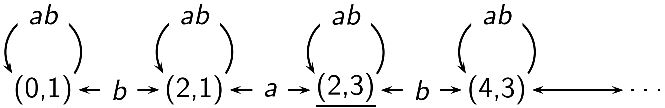
Consecutive numbers with lying



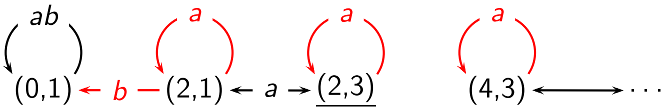
- Anne: "I know your number." Anne is lying



Consecutive numbers with lying

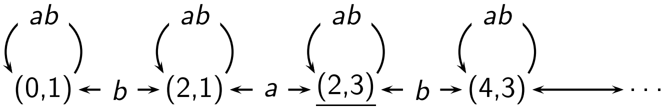


- Anne: "I know your number." Anne is lying

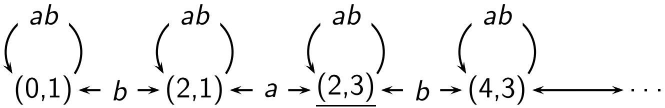


- Bill: "That's a lie."

Consecutive numbers with lying

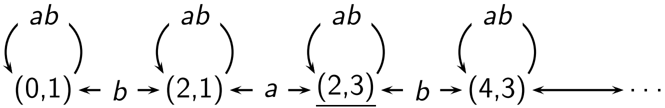


Consecutive numbers with lying

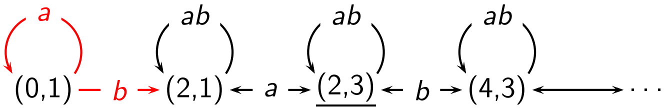


- Anne: "I do not know your number."

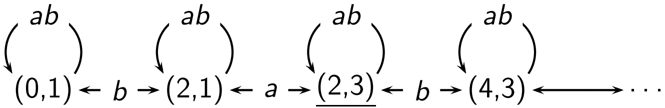
Consecutive numbers with lying



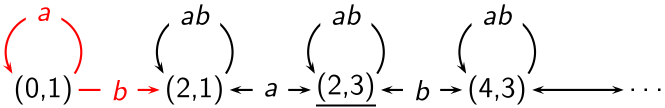
- Anne: "I do not know your number."



Consecutive numbers with lying

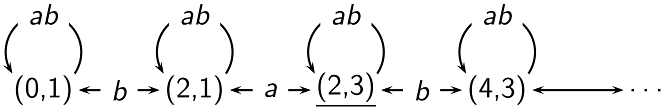


- Anne: "I do not know your number."

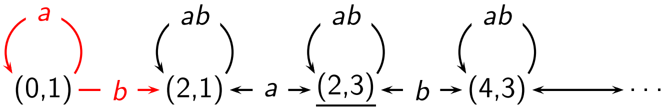


- Bill: "I know your number." Bill is lying

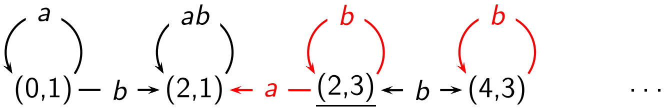
Consecutive numbers with lying



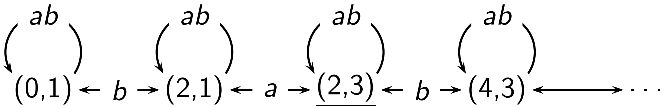
- Anne: "I do not know your number."



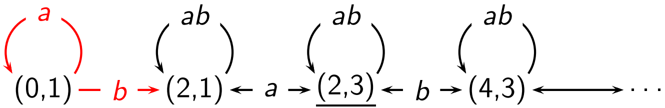
- Bill: "I know your number." Bill is lying



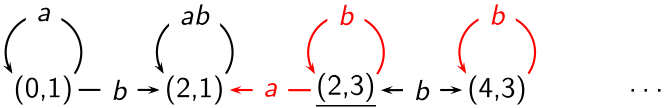
Consecutive numbers with lying



- Anne: “I do not know your number.”

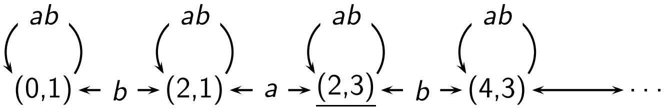


- Bill: “I know your number.” Bill is lying

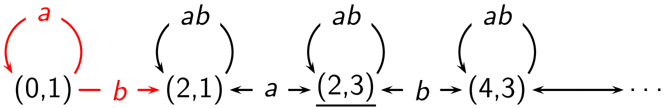


- Anne: “I know your number.” Anne is mistaken.

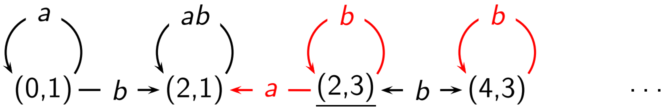
Consecutive numbers with lying



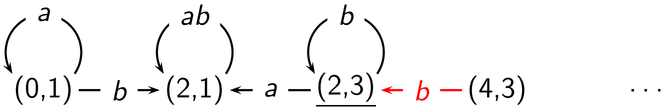
- Anne: “I do not know your number.”



- Bill: “I know your number.” Bill is lying



- Anne: “I know your number.” Anne is mistaken.



Results for lying announcement

- Public announcement and agent announcement can be represented by an *action model*, i.e., a general framework for epistemic dynamics. This translation provides axiomatizations (and upper bounds for complexity issues).
- A *skeptical* agent does not accept new information φ if it already believes $\neg\varphi$: more complex preconditions.
- Agents may distinguish between more and less plausible states, and more and less plausible actions: a truthful announcement is more plausible than a lying announcement (and a bluffing announcement in between the two). Unless you receive information to the contrary, you will assume the announcement is truthful. Otherwise, that it was bluffing. Otherwise, that it was a lie.
- In fair games players can distinguish lies from mistakes.

Let p stand for 'Oranges freeze in Sevilla' . . .

Oranges in Sevilla

p = Oranges freeze in Sevilla (as Hans claims)
a = speaker (me)
b = listener (you)

- Truthful announcement that p: if $B_a p$ and $!_a p$
- Lying announcement that p: $B_a \neg p$ and ¡$_a p$
- 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 $!_a p$
- The postcondition that holds: $B_b p$
- If you are skeptical, precondition: $\neg B_b \neg B_a p$

Further issues with lying

- Incorporating common knowledge / common belief:
  $B_a \neg\varphi \wedge \neg B_b \neg\varphi \wedge C_{ab}((B_a\varphi \vee B_a\neg\varphi) \wedge \neg(B_b\varphi \vee B_b\neg\varphi))$
- Insincere or strategic voting in social choice is a form of lying.
- Protocols with few liars or few lies. (Ulam Games)
- Signal analysis: noise versus intentional noise.
- Modelling a Liar Paradox in dynamic epistemic logic.
- Computational complexity of lying versus truthfulness
- Ref: Hans van Ditmarsch, *Dynamics of Lying*, Synthese, 2013.
  Contains references to other recent publications on lying.

Thank you!