



# LOGICAL DEFINITIONS OF LYING

嘘



Chiaki Sakama  
Wakayama University, Japan

*14<sup>th</sup> International Workshop on Trust in Agent Societies,  
Taipei, Taiwan, May 2, 2011*

# Trustful Agents and Lying

- Identifying **trustful** agents is vital in MAS.
- An agent is **trustful** only if he/she does not **lie**.
- Then, a question is: “**What is lying?**”

# Definition of Lying

- Studied by a number of philosophers and many different definitions exist.
- James E. Mahon (2008) provides a comprehensive study of lying.
- He examines 12 different definitions in the literature in an informal manner.

# Contributions

- Formulate 12 different definitions of lying using a multi-modal logic.
- Introduce conditions that should be satisfied by any logical definition of lying.
- Compare formal properties of each definition, and argue which definition is most intuitive.

# A Logic for Belief and Intention

- A propositional multimodal logic with 2 modalities:  $B_a\phi$  ( $a$  believes a sentence  $\phi$ ) and  $I_a\phi$  ( $a$  intends  $\phi$ ).
- The logic is an extension of  $KD45_n$  with the following axioms and inference rules.

(P) All propositional tautologies

(K<sub>B</sub>)  $B_a\phi \wedge B_a(\phi \supset \psi) \supset B_a\psi$     (K<sub>I</sub>)  $I_a\phi \wedge I_a(\phi \supset \psi) \supset I_a\psi$

(D<sub>B</sub>)  $B_a\phi \supset \neg B_a\neg\phi$     (D<sub>I</sub>)  $I_a\phi \supset \neg I_a\neg\phi$

(4<sub>B</sub>)  $B_a\phi \supset B_aB_a\phi$     (4<sub>IB</sub>)  $I_a\phi \supset B_aI_a\phi$

(5<sub>B</sub>)  $\neg B_a\phi \supset B_a\neg B_a\phi$     (5<sub>IB</sub>)  $\neg I_a\phi \supset B_a\neg I_a\phi$

(MP) from  $\vdash \phi$  and  $\vdash \phi \supset \psi$ , infer  $\vdash \psi$

(N<sub>B</sub>) from  $\vdash \phi$  infer  $\vdash B_a\phi$     (N<sub>I</sub>) from  $\vdash \phi$  infer  $\vdash I_a\phi$

# Semantics

- $B_a\phi$  (resp.  $I_a\phi$ ) holds iff  $\phi$  is true in all states of affairs compatible with  $a$ 's current beliefs (resp. intention).
- If  $\phi$  means that it rains,  $I_a\phi$  should be read as " $a$  intends to act in such a way that he/she brings about a state of affairs in which it rains".
- With this reading,  $(N_I)$  from  $\vdash \phi$  infer  $\vdash I_a\phi$  says that all theorems hold at all state of affairs that  $a$  might intend to bring about.

# Utterance

- A speech act of an agent is represented by  $\text{utter}_{ab}(\sigma)$  (an agent  $a$  utters a sentence  $\sigma$  to another agent  $b$ ), which satisfies the axioms:
- $(U_{IB}) \quad \text{utter}_{ab}(\sigma) \supset I_a(\text{utter}_{ab}(\sigma)) \wedge B_a(\text{utter}_{ab}(\sigma))$   
(if  $a$  utters  $\sigma$  to  $b$ , then  $a$  intends the act and is aware of the act.)
- $(U_{BB}) \quad \text{utter}_{ab}(\sigma) \supset B_b(\text{utter}_{ab}(\sigma)) \wedge B_a B_b(\text{utter}_{ab}(\sigma))$   
(if  $a$  utters  $\sigma$  to  $b$ , then  $b$  recognizes the act and  $a$  is aware of the recognition by  $b$ .)

# Four Necessary Conditions for Lying [Mahon 2008]

If a person lies, then

- the person makes a statement  
(statement condition)
- the person believes the statement to be false  
(untruthfulness condition)
- the untruthful statement is made to another person (addressee condition)
- the person intends that the other person's believing the untruthful statement to be true  
(intention to deceive addressee condition)



# Vrij's Definition

- (L1) To lie (to another person) is: to attempt to create a believed-false belief without forewarning (in another person) [Mahon 2008]
- $L1_{ab}(\sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge B_a B_b B_a \sigma$   
where  $a, b$  : agents,  $\sigma$  : sentence
- $a$  believes the falsity of  $\sigma$  ( $B_a \neg \sigma$ ), and intends to make  $b$  believe  $\sigma$  ( $I_a B_b \sigma$ ). Without forewarning,  $a$  believes that it is justified for  $b$  to believe that  $a$  believes  $\sigma$  ( $B_a B_b B_a \sigma$ ).
- L1 satisfies untruthfulness condition ( $B_a \neg \sigma$ ) and the intention to deceive addressee condition ( $I_a B_b \sigma$ ).

## Problems of (L1)

- (L1) does not satisfy the statement condition and addressee condition.
- (L1) requires no statement, so “feigning a yawn, wearing a hair-piece, making a phony smile, wearing an engagement ring when one is not engaged, ... is lying” [Mahon 2008].
- Mahon rejects (L1) because it is too broad as a definition of lying.

# OED Definition

- (L<sub>4</sub>) To lie (to another person) is: to make a false statement (to another person) with the intention to deceive (some person or other) [Mahon 2008]
- $L_{4ab}(\sigma, \lambda) = \text{utter}_{ab}(\sigma) \wedge \neg\sigma \wedge B_a B_b(\sigma \supset \lambda) \wedge B_a \neg\lambda \wedge I_a B_b \lambda$   
where  $a, b$  : agents,  $\sigma, \lambda$  : sentences
- $a$  utters a false statement ( $\text{utter}_{ab}(\sigma) \wedge \neg\sigma$ ), and  $a$  believes that  $b$  uses  $\sigma$  to reach a conclusion  $\lambda$  ( $B_a B_b(\sigma \supset \lambda)$ ).  $a$  believes the falsity of  $\lambda$  ( $B_a \neg\lambda$ ), and believing  $\lambda$  by  $b$  is what  $a$  intends to achieve ( $I_a B_b \lambda$ ).
- L<sub>4</sub> satisfies the statement/addressee condition ( $\text{utter}_{ab}(\sigma)$ ), and the intention to deceive addressee condition ( $I_a B_b \lambda$ ).

## Problems of (L4)

- (L<sub>4</sub>) does not satisfy untruthfulness condition ( $B_a \neg \sigma$ ), instead, it requires the statement to be false ( $\neg \sigma$ ) (falsity condition).
- “A person is to be judged as lying or not lying according to the intention of his own mind, not according to the truth or falsity of the matter itself.” — Saint Augustine, *Lying* (1952).
- Mahon rejects (L<sub>4</sub>) because “one can lie by being truthful with an intention to deceive, when it just so happens that one is mistaken”.

# Kupfer's Definition

- (L6) To lie (to another person) is: to make a believed-false statement (to another person) with the intention that that statement be believed to be true (by the other person) [Mahon 2008]
- $L6_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge B_a \neg \sigma \wedge I_a B_b \sigma$
- L6 satisfies the statement/addressee condition ( $\text{utter}_{ab}(\sigma)$ ), untruthfulness condition ( $B_a \neg \sigma$ ), and the intention to deceive addressee condition ( $I_a B_b \lambda$ ).
- (L6) is considered a “standard” definition of lying [Mahon 2008].

## Problems of (L6)

- Lying involves two distinct intentions to deceive [Frankfurt 1992]:
  - (i) intention that the hearer  $b$  believes a believed-false statement  $\sigma$ , and
  - (ii) intention that the hearer  $b$  believes that the speaker  $a$  believes the statement  $\sigma$ .
- (L6) fails to represent (ii), that is, an intention to deceive about one's belief in the truth of the statement one makes (**believed truthfulness condition**) [Mahon 2008].

# Mahon's Definition

- (L6\*) To lie (to another person) is: to make a believed-false statement (to another person) either with the intention that that statement be believed to be true (by the other person), or with the intention that it be believed (by the other person) that that statement is believed to be true (by the person making the statement), or with both intentions [Mahon 2008] .
- $L6^*_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge B_a \neg \sigma \wedge (I_a B_b \sigma \vee I_a B_b B_a \sigma)$
- L6\* represents the believed truthfulness condition by  $(I_a B_b B_a \sigma)$ .
- Mahon asserts that (L6) and (L6\*) are two best definitions of lying.

# Other Definitions

- $L2_{ab}(\sigma) = (\text{utter}_{aa}(\sigma) \vee \text{utter}_{ab}(\sigma)) \wedge B_a \neg \sigma$
- $L3_{ab}(\sigma, \lambda) = \text{utter}_{ab}(\sigma) \wedge B_a B_b (\sigma \supset \lambda) \wedge B_a \neg \lambda \wedge I_a B_b \lambda$
- $L5_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge B_a \neg \sigma \wedge \neg \sigma \wedge I_a B_b \sigma$
- $L7_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge B_a \neg \sigma \wedge I_a B_b \sigma \wedge I_a B_b B_a \sigma$
- $L8_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge (\neg B_a \sigma \vee B_a \neg \sigma) \wedge B_a B_b B_a \sigma$   
 $\wedge B_a B_b I_a B_b B_a \sigma$
- $L9_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge B_a \neg \sigma \wedge I_a B_b \sigma \wedge I_a B_b B_a \sigma$   
 $\wedge I_a B_b I_a B_b B_a \sigma$
- $L10_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge \neg B_a \sigma \wedge \neg \sigma \wedge B_b B_a \sigma$   
 $\wedge \neg B_a \neg B_b B_a \sigma$
- $L11_{ab}(\sigma) = \text{utter}_{ab}(\sigma) \wedge B_a \neg \sigma$   
 $\wedge B_a B_b (\text{utter}_{ab}(\sigma) \supset \neg B_a \neg \sigma)$



# Five Conditions

- Lying on valid sentences is impossible (inability to lie on valid sentences, or inability-T).
- Lying on contradictory sentences is impossible (inability to lie on contradictory sentences, or inability- $\perp$ ).
- Lying on two sentences  $\sigma$  and  $\neg\sigma$  at the same time is impossible (inability to lie on mutually conflicting sentences, or inability- $\neg$ ).
- A liar is aware of his/her dishonest act (awareness).
- Lying to oneself leads to contradiction (self-contradiction).

# Formal Properties

Let  $LIE_{ab}(\sigma)$  be one of the definitions of (L1), (L2), and (L5)-(L11).

- **(inability-T)**  $\vdash LIE_{ab}(T) \supset \perp$  holds if  $LIE_{ab}(\sigma)$  includes either  $B_a \neg \sigma$  or  $B_a \neg \sigma$  or  $\neg \sigma$ .
- **(inability- $\perp$ )**  $\vdash LIE_{ab}(\perp) \supset \perp$  holds if  $LIE_{ab}(\sigma)$  includes either  $I_a B_b \sigma$  or  $I_a B_b B_a \sigma$  or  $B_a B_b B_a \sigma$ .
- **(inability- $\neg$ )**  $\vdash LIE_{ab}(\sigma) \wedge LIE_{ab}(\neg \sigma) \supset \perp$  holds if  $LIE_{ab}(\sigma)$  includes either  $B_a \neg \sigma$  or  $\neg \sigma$ .

# Formal Properties

- **(awareness)**  $\vdash \text{LIE}_{ab}(\sigma) \supset B_a(\text{LIE}_{ab}(\sigma))$   
holds for any sentence  $\sigma$  only if  $\text{LIE}_{ab}(\sigma)$  does not include the falsity condition  $\neg\sigma$ .
- **(self-contradiction)**  $\vdash \text{LIE}_{aa}(\sigma) \supset \perp$   
holds for any sentence  $\sigma$  if  $\text{LIE}_{ab}(\sigma)$  includes either  $B_a\neg\sigma \wedge I_a B_b \sigma$  or  $B_a\neg\sigma \wedge I_a B_b B_a \sigma$  or  $B_a\neg\sigma \wedge B_b B_a \sigma$ .
- The results hold for (L<sub>3</sub>) and (L<sub>4</sub>) that have the additional parameter  $\lambda$ .

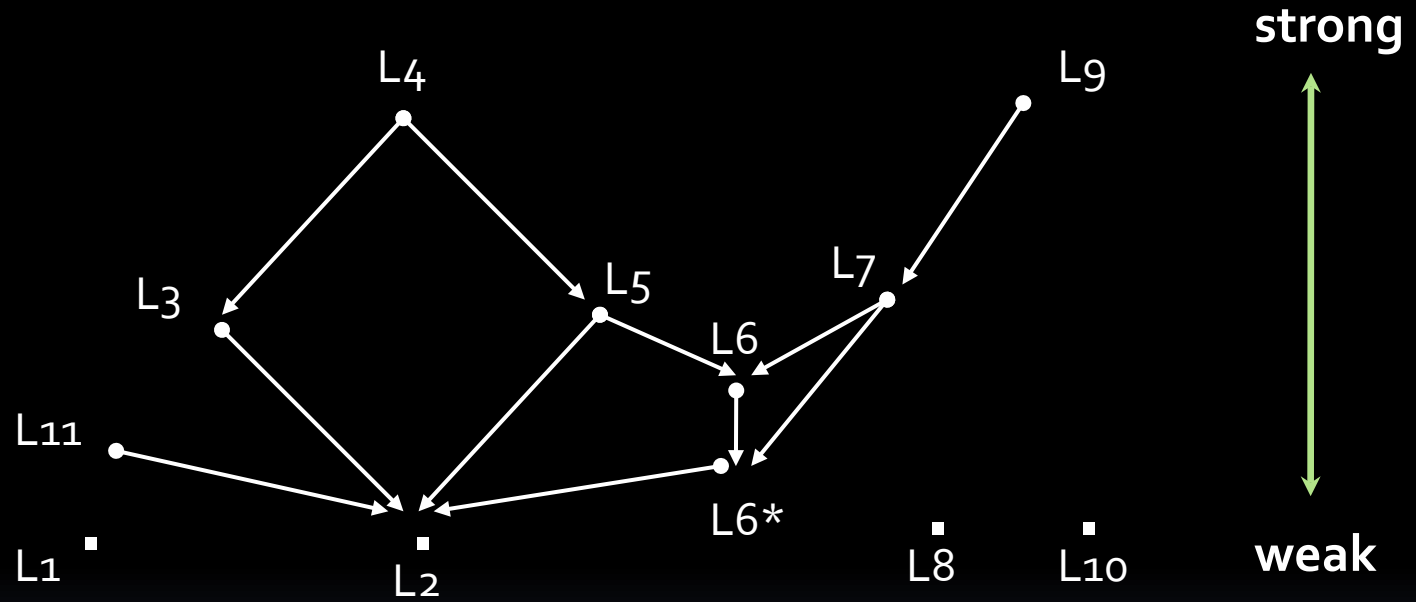
# Comparison of 12 Definitions

	L1	L2	L3	L4	L5	L6	L6*	L7	L8	L9	L10	L11
statement		x	x	x	x	x	x	x	x	x	x	x
addressee		(x)	x	x	x	x	x	x	x	x	x	x
untruthful	x	x			x	x	x	x	(x)	x		x
intention	x		x	x	x	x	(x)	x		x		
believed truthful							(x)	x		x		
falsity				x	x						x	
inability-T	x	x		x	x	x	x	x	x	x	x	x
inability- $\perp$	x				x	x	x	x	x	x	x	x
inability- $\neg$	x	x		x	x	x	x	x		x	x	x
awareness	x	x	x			x	x	x	x	x		x
self-contradiction	x		x	x	x	x	x	x	x	x	x	x

necessary conditions by Mahon ; conditions considered in this paper

(x) means that the condition is included as a disjunct

# Relationship between 12 Definitions



# Conclusion

- Logical definitions of lying are provided, and their formal properties are analyzed.
- The results provide formal accounts for Mahon's informal arguments.
- The 5 new conditions that logically justify the act of lying, together with 4 necessary conditions by Mahon that empirically support lying, serve as criteria for judging whether yet another definition of lying is appropriate or not.

# Related Studies by the Author

- Chiaki Sakama, Martin Caminada, and Andreas Herzig: **A Logical Account of Lying**. 12<sup>th</sup> European Conf. on Logics in AI (JELIA), LNAI 6341, Springer, 2010.
- Chiaki Sakama and Martin Caminada : **The Many Faces of Deception**. Thirty Years of Nonmonotonic Reasoning (NonMon@30), Lexington, KY, USA, October 2010.

# Forthcoming

- Ngoc-Hieu Nguyen, Tran Son, Enrico Pontelli, and Chiaki Sakama: **ASP-Prolog for Negotiation Among Dishonest Agents**. 11<sup>th</sup> Int'l Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR), Vancouver, May 2011.
- Chiaki Sakama, Tran Son, and Enrico Pontelli : **A Logical Formulation for Negotiation Among Dishonest Agents**. IJCAI-2011, Barcelona, July 2011.
- Chiaki Sakama : **Dishonest Reasoning by Abduction**. IJCAI-2011, Barcelona, July 2011.