

# THE MANY FACES OF DECEPTION

Chiaki Sakama

(Wakayama University, Japan)

Martin Caminada

(University of Luxembourg, Luxembourg)



*Thirty Years of Nonmonotonic Reasoning, Lexington, KY, USA; October 2010  
Revised, March 2011*

# Deception

- **Deception** is an act whereby one person causes another person to have a false belief.
- In spite of its commonality in human society, little study has been devoted to developing a formal account of deception.
- By understanding what deception is, one can consider the best ways of using deception to achieve a particular goal, and the best ways in which one could avoid being deceived.

# Definition of Deception

- Studied by a number of philosophers and many different definitions exist.
- Chisholm and Feehan (1977) provide eight basic ways of deception.
- These categories are further divided depending on whether the deception is intended or not.

# Contributions

- Formulate 8 different categories of deception using a modal logic of belief and action.
- Investigate formal properties of deception and propose postulates that should ideally be satisfied by agents.
- Argue the difference between intended deception, lying and withholding information.

# Logic for Belief and Action

[Pörn 1977,1989]

- A multi-modal logic with 3 modalities
  - $B_a\phi$ :  $a$  believes  $\phi$
  - $D_a\phi$ : it is necessary for something which  $a$  does  $\phi$
  - $D'_a\phi$ : but for  $a$ 's action it would be the case that  $\phi$
- A unary predicate representing a speech act
  - $utter_{ab}(\phi)$ :  $a$  expresses a sentence  $\phi$  to  $b$

# The Axiom System for BDD'

- The axiom system for the B-operator is **KD45**, while those for the D-operator and the D'-operator are **KT** and **KD**, respectively.
- A logic **BDD'** has the following axioms and inference rules.

(P) All propositional tautologies

(U<sub>c</sub>)  $\text{utter}_{ab}(\phi \wedge \psi) \equiv \text{utter}_{ab}(\phi) \wedge \text{utter}_{ab}(\psi)$

(K<sub>B</sub>)  $B_a\phi \wedge B_a(\phi \supset \psi) \supset B_a\psi$  (K<sub>D</sub>)  $D_a\phi \wedge D_a(\phi \supset \psi) \supset D_a\psi$

(K<sub>D'</sub>)  $D'_a\phi \wedge D'_a(\phi \supset \psi) \supset D'_a\psi$  (T<sub>D</sub>)  $D_a\phi \supset \phi$

(D<sub>B</sub>)  $B_a\phi \supset \neg B_a\neg\phi$  (D<sub>D'</sub>)  $D'_a\phi \supset \neg D'_a\neg\phi$

(4<sub>B</sub>)  $B_a\phi \supset B_aB_a\phi$  (5<sub>B</sub>)  $\neg B_a\phi \supset B_a\neg B_a\phi$

(MP)  $\phi, \phi \supset \psi / \psi$  (N<sub>B</sub>)  $\phi / B_a\phi$

(N<sub>D</sub>)  $\phi / D_a\phi$  (N<sub>D'</sub>)  $\phi / D'_a\phi$

# Two Action Operators

- $E_a\phi = D_a\phi \wedge \neg D'_a\phi$

: it is necessary for something which  $a$  does  $\phi$ ,  
but for  $a$ 's action it would not be the case that  $\phi$ .  
"a brings it about that  $\phi$ "

- $F_a\phi = \neg D_a\neg\phi \wedge \neg D'_a\phi$

: it is compatible with everything which  $a$  does  
that  $\phi$ , but for  $a$ 's action it would not be the case  
that  $\phi$ .

"a let it be the case that  $\phi$ "

# Properties of E and F

- $(T_E) E_a\phi \supset \phi$
- $(D_E) E_a\phi \supset \neg E_a\neg\phi$
- $(K_E) E_a(\phi \supset \psi) \supset (E_a\phi \supset E_a\psi)$
- $(C_E) (E_a\phi \wedge E_a\psi) \supset E_a(\phi \wedge \psi)$
- $(No) \neg E_aT \wedge \neg F_aT$
- $(EF) E_a\phi \supset F_a\phi$

✓  $(T_E)$  represents the success of actions, while  $(No)$  represents that no agent can bring about what is logically true.



# Causal Relation

[Sandu 1986]

- $\phi \Rightarrow \psi$ : “ $\psi$  is a consequence of  $\phi$ ”
- It satisfies the following axioms and inference rules:
  - (C1)  $\neg(\phi \Rightarrow \perp)$
  - (C2)  $(\phi \Rightarrow \psi \wedge \psi \Rightarrow \chi) \supset \phi \Rightarrow \chi$
  - (C3)  $(\phi \Rightarrow \psi) \supset \neg(\psi \Rightarrow \phi)$
  - (C4)  $\neg(\phi \Rightarrow \psi \wedge \neg\phi \Rightarrow \psi)$
  - (C5)  $(\phi \Rightarrow \psi) \supset (\phi \wedge \psi)$
  - (C6)  $(\phi \Rightarrow \psi \wedge \phi \Rightarrow \chi) \supset (\phi \Rightarrow \psi \wedge \chi)$
  - (C7)  $\phi \equiv \psi / (\phi \Rightarrow \chi) \equiv (\psi \Rightarrow \chi)$
  - (C8)  $\phi \equiv \psi / (\chi \Rightarrow \phi) \equiv (\chi \Rightarrow \psi)$
- A logic **BDD'C** is an extension of BDD' with (C1)-(C8).

# Deception by Commission

[Chisholm and Feehan, 1977]

- $a, b$ : two agents;  $p$ : a false proposition
- (DC<sub>1</sub>)  $a$  contributes causally toward  $b$ 's acquiring the belief in  $p$  (positive deception simpliciter).
- (DC<sub>2</sub>)  $a$  contributes causally toward  $b$ 's continuing in the belief in  $p$  (positive deception secundum quid).
- (DC<sub>3</sub>)  $a$  contributes causally toward  $b$ 's ceasing to believe in  $\neg p$  (negative deception simpliciter).
- (DC<sub>4</sub>)  $a$  contributes causally toward preventing  $b$  from acquiring the belief in  $\neg p$  (negative deception secundum quid).

# Positive Deception by Commission: Definition

- $\text{PDSC}_{ab}(\sigma) = B_a \neg \sigma \wedge (\text{utter}_{ab}(\sigma) \Rightarrow E_a B_b \sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by PDSC if  $a$  utters a believed-false sentence  $\sigma$  and the utterance causally brings it about that  $b$  believes  $\sigma$ .
- $\text{PDSQC}_{ab}(\sigma) = B_a \neg \sigma \wedge (\text{utter}_{ab}(\sigma) \Rightarrow F_a B_b \sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by PDSQC if  $a$  utters a believed-false sentence  $\sigma$  and the utterance causally lets it be the case that  $b$  believes  $\sigma$ .

# Example

- Suppose a salesperson  $a$  who believes that a product has no value ( $B_a \neg \text{valuable}$ ).
- If the salesperson utters to a customer  $b$  that it is valuable ( $\text{utter}_{ab}(\text{valuable})$ )
- and the speech act leads the customer, who disbelieves that the product is valuable, to believing it valuable ( $E_a B_b \text{valuable}$ ), then it is  $\text{PDSC}_{ab}(\text{valuable})$ .
- On the other hand, if the speech act could let the customer, who believes that the product is valuable, continue to have the (wrong) belief ( $F_a B_b \text{valuable}$ ), then it is  $\text{PDSQC}_{ab}(\text{valuable})$ .

## Negative Deception by Commission: Definition

- $\text{NDSC}_{ab}(\sigma) = B_a \neg \sigma \wedge (\text{utter}_{ab}(\sigma) \Rightarrow E_a \neg B_b \neg \sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by NDSC if  $a$  utters a believed-false sentence  $\sigma$  and the utterance causally brings it about that  $b$  disbelieves  $\neg \sigma$ .
- $\text{NDSQC}_{ab}(\sigma) = B_a \neg \sigma \wedge (\text{utter}_{ab}(\sigma) \Rightarrow F_a \neg B_b \neg \sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by NDSQC if  $a$  utters a believed-false sentence  $\sigma$  and the utterance causally lets it be the case that  $b$  disbelieves  $\neg \sigma$  (thus preventing  $b$  from acquiring  $\neg \sigma$ ).

# Example

- Consider again a salesperson  $a$  who believes that a product has no value ( $B_a \neg \text{valuable}$ ).
- If the salesperson utters to a customer  $b$  that it is valuable ( $\text{utter}_{ab}(\text{valuable})$ )
- and the speech act makes the customer believe the possibility of the value ( $E_a \neg B_b \neg \text{valuable}$ ), then it is  $\text{NDSC}_{ab}(\text{valuable})$ .
- On the other hand, if by the speech act the customer continues to believe the possibility of the value ( $F_a \neg B_b \neg \text{valuable}$ ), then it is  $\text{NDSQC}_{ab}(\text{valuable})$ .

# Deception by Omission

[Chisholm and Feehan, 1977]

- $a, b$ : two agents;  $p$ : a false proposition
- (DO<sub>1</sub>)  $a$  allows  $b$  to acquire the belief in  $p$   
(positive deception simpliciter).
- (DO<sub>2</sub>)  $a$  allows  $b$  to continue in the belief in  $p$   
(positive deception secundum quid).
- (DO<sub>3</sub>)  $a$  allows  $b$  to cease to have the belief in  $\neg p$   
(negative deception simpliciter).
- (DO<sub>4</sub>)  $a$  allows  $b$  to continue without the belief in  $\neg p$   
(negative deception secundum quid).

# Remark

- Here, an agent **allows** a certain state of affairs to occur provided only (i) he/she could prevent that state of affairs from occurring, and (ii) he/she does not thus prevent it from occurring.
- We capture the act that "*a* allows *b*" as "*a* makes no utterance to *b* on a believed-true sentence".
- That is, *a* proactively do nothing to affect *b*'s state of mind of believing what is true.



## Positive Deception by Omission: Definition

- $\text{PDSO}_{ab}(\sigma) = B_a \neg\sigma \wedge (\neg \text{utter}_{ab}(\neg\sigma) \Rightarrow E_a B_b \sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by PDSO if  $a$  does not utter a believed-true sentence  $\neg\sigma$  and the non-utterance causally brings it about that  $b$  believes  $\sigma$ .
- $\text{PDSQO}_{ab}(\sigma) = B_a \neg\sigma \wedge (\neg \text{utter}_{ab}(\neg\sigma) \Rightarrow F_a B_b \sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by PDSQO if  $a$  does not utter a believed-true sentence  $\neg\sigma$  and the non-utterance causally lets it be the case that  $b$  believes  $\sigma$ .

# Example

- Suppose a child  $a$  who believes that he/she failed to get a passing grade ( $B_a \neg \text{pass}$ ), but does not utter the fact to his/her parent  $b$  ( $\neg \text{utter}_{ab}(\neg \text{pass})$ ).
- If the non-utterance leads the parent to believing that the child gets a passing grade ( $E_a B_b \text{pass}$ ), then it is  $\text{PDSO}_{ab}(\text{pass})$ .
- On the other hand, if the non-utterance leads the parent to retaining the wrong belief ( $F_a B_b \text{pass}$ ), then it is  $\text{PDSQO}_{ab}(\text{pass})$ .

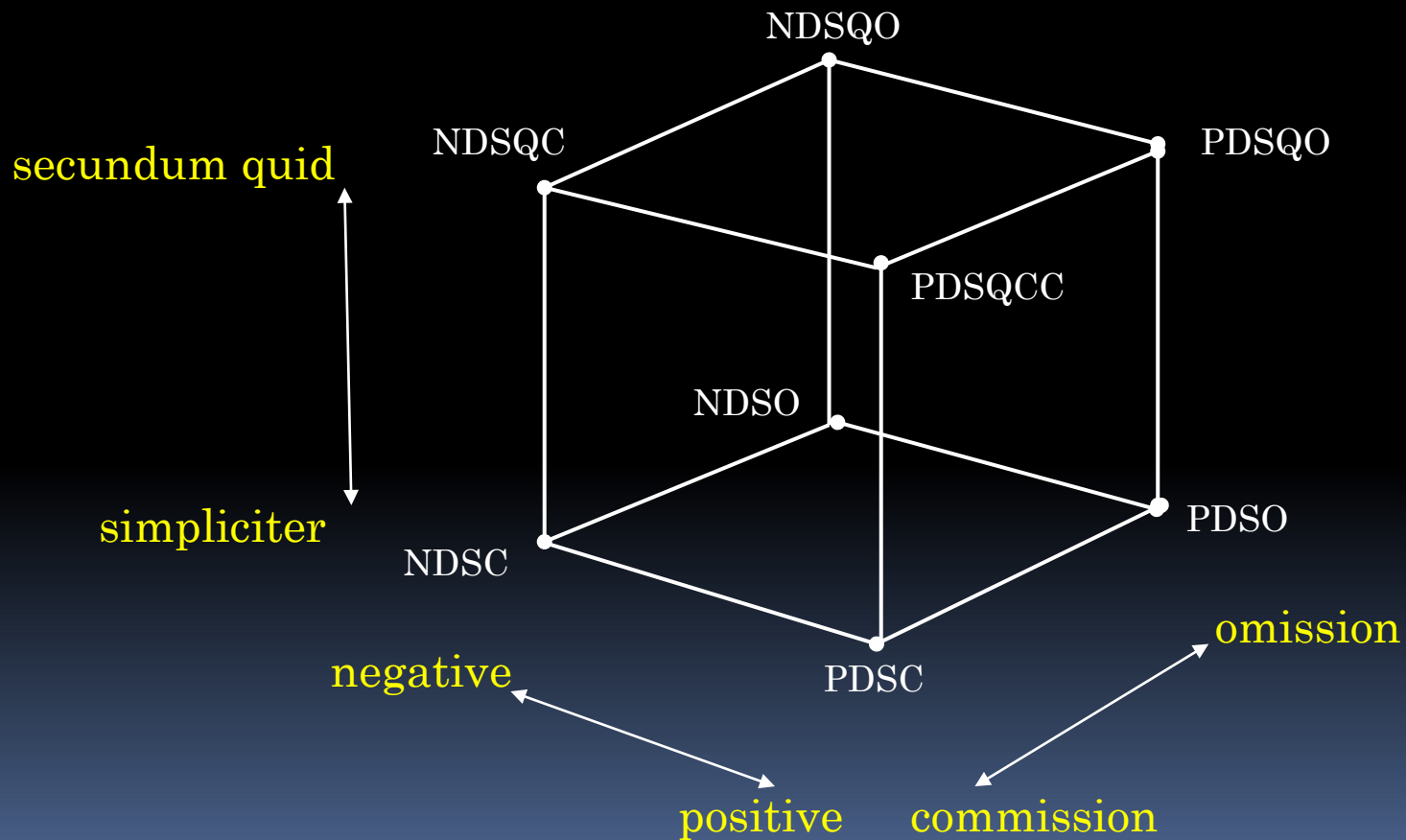
# Negative Deception by Omission: Definition

- $\text{NDSO}_{ab}(\sigma) = B_a \neg\sigma \wedge (\neg \text{utter}_{ab}(\neg\sigma) \Rightarrow E_a \neg B_b \neg\sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by NDSO if  $a$  does not utter a believed-true sentence  $\neg\sigma$  and the non-utterance causally brings it about that  $b$  disbelieves  $\neg\sigma$ .
- $\text{NDSQO}_{ab}(\sigma) = B_a \neg\sigma \wedge (\neg \text{utter}_{ab}(\neg\sigma) \Rightarrow F_a \neg B_b \neg\sigma)$   
:  $a$  deceives  $b$  on  $\sigma$  by NDSQO if  $a$  does not utter a believed-true sentence  $\neg\sigma$  and the non-utterance causally lets it be the case that  $b$  disbelieves  $\neg\sigma$ .

# Example

- Consider again a child  $a$  who believes that he/she failed to get a passing grade ( $B_a \neg \text{pass}$ ), but does not utter the fact to his/her parent  $b$  ( $\neg \text{utter}_{ab}(\neg \text{pass})$ ).
- If the non-utterance makes the parent believe the possibility of the child's getting a passing grade ( $E_a \neg B_b \neg \text{pass}$ ), then it is  $\text{NDSO}_{ab}(\text{pass})$ .
- On the other hand, if the non-utterance leads the parent to retaining the belief on the possibility of the child's getting a passing grade, ( $F_a \neg B_b \neg \text{pass}$ ), then it is  $\text{NDSQO}_{ab}(\text{pass})$ .

# 8 Categories of Deception



# Deception by Commission (DC) v.s. Deception by Omission (DO)

- DC is considered to be **more sinful** than DO because in DC a deceiver behaves more proactively than in DO.
- Thus, PDSC is worse (or more sinful) than PDSO, PDSQC is worse than PDSQO, NDSC is worse than NDSO, and NDSQC is worse than NDSQO.
- These relations imply a guideline for a speech-act of deception that should ideally be satisfied by the agents, for moral as well as for self-interested reasons (lower punishments if caught).

# Postulates for DC and DO

(PDSC-PDSO)

$$B_a(\text{PDSC}_{ab}(\sigma)) \wedge B_a(\text{PDSO}_{ab}(\sigma)) \supset \neg \text{PDSC}_{ab}(\sigma)$$

“if  $a$  believes that both PDSC and PDSO are effective, then  $a$  does not opt for PDSC.”

(PDSQC-PDSQO)

$$B_a(\text{PDSQC}_{ab}(\sigma)) \wedge B_a(\text{PDSQO}_{ab}(\sigma)) \supset \neg \text{PDSQC}_{ab}(\sigma)$$

(NDSC-NDSO)

$$B_a(\text{NDSC}_{ab}(\sigma)) \wedge B_a(\text{NDSO}_{ab}(\sigma)) \supset \neg \text{NDSC}_{ab}(\sigma)$$

(NDSQC-NDSQO)

$$B_a(\text{NDSQC}_{ab}(\sigma)) \wedge B_a(\text{NDSQO}_{ab}(\sigma)) \supset \neg \text{NDSQC}_{ab}(\sigma)$$

# Formal Properties

- Let  $Dec$  be one of the 8 categories of deception. Then the followings are proved.
- It is inconsistent to deceive on  $valid(T)$  or  $contradictory(\perp)$  sentences.  
 $Dec_{ab}(T) \supset \perp$  and  $Dec_{ab}(\perp) \supset \perp$
- Self-deception leads to contradiction.  
 $Dec_{aa}(\sigma) \supset \perp$
- Deceptions do not work conjunctively.  
 $Dec_{ab}(\sigma) \wedge Dec_{ab}(\lambda) \not\supset Dec_{ab}(\sigma \wedge \lambda)$



# Intended Deception

- Deception is often accompanied by **intention**.
- The logic BDD'C is extended to **BDD'CI** by introducing the modality  $I_a\phi$  representing that "an agent  $a$  intends  $\phi$ ".
- The I-operator follows the KD45 system.

(P) All propositional tautologies

(K<sub>I</sub>)  $I_a\phi \wedge I_a(\phi \supset \psi) \supset I_a\psi$     (D<sub>I</sub>)  $I_a\phi \supset \neg I_a\neg\phi$

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

(MP)  $\phi, \phi \supset \psi / \psi$     (N<sub>I</sub>)  $\phi / I_a\phi$

# Intended Deception by Commission: Definition

- $$\text{IPDSC}_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a B_b \phi \wedge B_a B_b (\sigma \supset \phi) \\ \wedge (\text{utter}_{ab}(\sigma) \Rightarrow E_a B_b \phi)$$

:  $a$  deceives  $b$  on  $\sigma$  for  $\phi$  by IPDSC if

- $a$  has an intention to make  $b$  believe the believed-false sentence  $\phi$ ,
- $a$  believes that a sentence  $\sigma$  leads  $b$  to believing  $\phi$ , and
- the utterance of  $\sigma$  brings it about that  $b$  believes  $\phi$ .

- $$\text{IPDSQC}_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a B_b \phi \wedge B_a B_b (\sigma \supset \phi) \\ \wedge (\text{utter}_{ab}(\sigma) \Rightarrow F_a B_b \phi)$$

- $$\text{INDSC}_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a \neg B_b \neg \phi \wedge B_a B_b (\sigma \supset \phi) \\ \wedge (\text{utter}_{ab}(\sigma) \Rightarrow E_a \neg B_b \neg \phi)$$

- $$\text{INDSQC}_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a \neg B_b \neg \phi \wedge B_a B_b (\sigma \supset \phi) \\ \wedge (\text{utter}_{ab}(\sigma) \Rightarrow F_a \neg B_b \neg \phi)$$

# Example

- Bob and Mary are working at the same office and Bob is playing a computer game. They know that their boss usually arrives at the office at ten o'clock ( $B_a B_b (\text{ten} \supset \text{boss})$ ).
- Mary knows that the boss will not arrive at ten today for some reasons ( $B_a \neg \text{boss}$ ), but intends Bob to believe he'll arrive ( $I_a B_b \text{boss}$ ), in order for Bob to stop playing the game.
- When the clock strikes ten, she utters "Now it's ten o'clock" ( $\text{utter}_{ab}(\text{ten})$ ).
- From this utterance, it follows that Bob realizes the boss is coming ( $E_a B_b \text{boss}$ ).
- This is an example of  $\text{IPDSC}_{ab}(\text{ten}, \text{boss})$ .

# DC v.s. Intended DC (IDC)

When  $\phi \equiv \sigma$ , IDC is simplified as

- $IPDSC_{ab}(\sigma, \sigma) = PDSC_{ab}(\sigma) \wedge I_a B_b \sigma$
- $IPDSQC_{ab}(\sigma, \sigma) = PDSQC_{ab}(\sigma) \wedge I_a B_b \sigma$
- $INDSC_{ab}(\sigma, \sigma) = NDSC_{ab}(\sigma) \wedge I_a \neg B_b \neg \sigma$
- $INDSQC_{ab}(\sigma, \sigma) = NDSQC_{ab}(\sigma) \wedge I_a \neg B_b \neg \sigma$

# Intended Deception by Omission: Definition

- $IPDSO_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a B_b \phi$   
 $\wedge B_a B_b (\neg \sigma \supset \neg \phi) \wedge (\neg \text{utter}_{ab}(\neg \sigma) \Rightarrow E_a B_b \phi)$
- $IPDSQO_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a B_b \phi$   
 $\wedge B_a B_b (\neg \sigma \supset \neg \phi) \wedge (\neg \text{utter}_{ab}(\neg \sigma) \Rightarrow F_a B_b \phi)$
- $INDSO_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a \neg B_b \neg \phi$   
 $\wedge B_a B_b (\neg \sigma \supset \neg \phi) \wedge (\neg \text{utter}_{ab}(\neg \sigma) \Rightarrow E_a \neg B_b \neg \phi)$
- $INDSQO_{ab}(\sigma, \phi) = B_a \neg \phi \wedge I_a \neg B_b \neg \phi$   
 $\wedge B_a B_b (\neg \sigma \supset \neg \phi) \wedge (\neg \text{utter}_{ab}(\neg \sigma) \Rightarrow F_a \neg B_b \neg \phi)$

# DO v.s. Intended DO (IDO)

When  $\phi \equiv \sigma$ , IDO is simplified as

- $IPDSO_{ab}(\sigma, \sigma) = PDSO_{ab}(\sigma) \wedge I_a B_b \sigma$
- $IPDSQO_{ab}(\sigma, \sigma) = PDSQO_{ab}(\sigma) \wedge I_a B_b \sigma$
- $INDSO_{ab}(\sigma, \sigma) = NDSO_{ab}(\sigma) \wedge I_a \neg B_b \neg \sigma$
- $INDSQO_{ab}(\sigma, \sigma) = NDSQO_{ab}(\sigma) \wedge I_a \neg B_b \neg \sigma$

# Postulates for IDC and IDO

(IPDSC-IPDSO)

$$B_a(\text{IPDSC}_{ab}(\sigma, \phi)) \wedge B_a(\text{IPDSO}_{ab}(\sigma, \phi)) \supset \neg \text{IPDSC}_{ab}(\sigma, \phi)$$

(IPDSQC-IPDSQO)

$$B_a(\text{IPDSQC}_{ab}(\sigma, \phi)) \wedge B_a(\text{IPDSQO}_{ab}(\sigma, \phi)) \\ \supset \neg \text{IPDSQC}_{ab}(\sigma, \phi)$$

(INDSC-INDSO)

$$B_a(\text{INDSC}_{ab}(\sigma, \phi)) \wedge B_a(\text{INDSO}_{ab}(\sigma, \phi)) \supset \neg \text{INDSC}_{ab}(\sigma, \phi)$$

(INDSQC-INDSQO)

$$B_a(\text{INDSQC}_{ab}(\sigma, \phi)) \wedge B_a(\text{INDSQO}_{ab}(\sigma, \phi)) \\ \supset \neg \text{INDSQC}_{ab}(\sigma, \phi)$$

# Three Types of Speech Acts

(John Austin: “How to do things with words”. 1962)

- A **locutionary act** is the act of saying something. (ex. utterance)
- An **illocutionary act** is the act made in saying something. (ex. promise)
- A **perlocutionary act** is the act made by saying something. (ex. persuasion)



# Lying

- Deception is a **perlocutionary act** that involves a success or an achievement of the act.
- By contrast, **lying** is *not* a perlocutionary act; whether or not an act of lying has occurred does not depend on whether or not a particular effect has been produced in another.
- Thus, **deceiving differs from lying**.

# Lie v.s. Deception

- In “A logical account of lying” by Sakama, Caminada, and Herzig (JELIA 2010), lies are defined as

$$\text{LIE}_{ab}(\sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \text{utter}_{ab}(\sigma)$$

: “*a* lies to *b* on a sentence  $\sigma$  if *a* utters the believed-false sentence  $\sigma$  to *b* with the intention that  $\sigma$  is believed by *b*.”

# Lie v.s. Deception

- Comparing

$$\text{LIE}_{ab}(\sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \text{utter}_{ab}(\sigma)$$

$$\text{IPDSC}_{ab}(\sigma, \sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \wedge (\text{utter}_{ab}(\sigma) \Rightarrow E_a B_b \sigma)$$

$$\text{IPDSQC}_{ab}(\sigma, \sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \wedge (\text{utter}_{ab}(\sigma) \Rightarrow F_a B_b \sigma)$$

the following implications hold:

- $\text{IPDSC}_{ab}(\sigma, \sigma) \supset \text{LIE}_{ab}(\sigma)$
- $\text{IPDSQC}_{ab}(\sigma, \sigma) \supset \text{LIE}_{ab}(\sigma)$

# Withholding Information

- **Withholding information** is to fail to offer information that would help someone acquire true beliefs and/or correct false beliefs.
- Not all cases of withholding information constitute deception.

# WI v.s. Deception

- Withholding information (WI) is defined as

$$WI_{ab}(\sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \neg \text{utter}_{ab}(\neg \sigma)$$

: “*a* makes no utterance to *b* on a believed-true sentence  $\neg \sigma$  with the intention that  $\sigma$  is believed by *b*.”

# WI v.s. Deception

- Comparing

$$WI_{ab}(\sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \neg \text{utter}_{ab}(\neg \sigma)$$

$$IPDSO_{ab}(\sigma, \sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \wedge (\neg \text{utter}_{ab}(\neg \sigma) \Rightarrow E_a B_b \sigma)$$

$$IPDSQO_{ab}(\sigma, \sigma) = B_a \neg \sigma \wedge I_a B_b \sigma \wedge \wedge (\neg \text{utter}_{ab}(\neg \sigma) \Rightarrow F_a B_b \sigma)$$

the following implications hold:

- $IPDSO_{ab}(\sigma, \sigma) \supset WI_{ab}(\sigma)$
- $IPDSQO_{ab}(\sigma, \sigma) \supset WI_{ab}(\sigma)$

# Action Logic is Nonmonotonic

- Action logic is **nonmonotonic** in the sense that an agent does not necessarily bring about all the consequences of his/her actions.
- So the following inference rules do **not** hold in general:

$$\frac{\phi \supset \psi}{E_a \phi \supset E_a \psi}$$

$$\frac{\phi \supset \psi}{F_a \phi \supset F_a \psi}$$

# Nonmonotonicity in Deception

- Deception based on action logic is thereby nonmonotonic. For instance,  $PDSC_{ab}(\sigma)$  implies  $E_a B_b \sigma$ , but  $E_a B_b \sigma$  and  $B_a \sigma \supset \neg B_a \neg \sigma$  do not imply  $E_a \neg B_a \neg \sigma$ .
- By this fact,  $PDSC_{ab}(\sigma)$  does not imply  $NDSC_{ab}(\sigma)$ .
- Positive deception does not imply negative deception in general.



# Nonmonotonicity in Deception

- “Truth, lies and bullshit, distinguishing classes of dishonesty” by Caminada (SS@IJCAI,2009) studies another type of deception introduced by Adler (1997).
- A deceiver asserts what he/she believes true, while, at the same time, he/she conceals something of the truth hoping that a hearer will make an incorrect inference based on incomplete beliefs.
- For instance, if the speaker tells the believed-true statement that *Tweety is a bird*, while withholding the fact that *Tweety is a penguin*, then the hearer might yield the default conclusion that *Tweety flies*, which the speaker believes to be false.

# Attempted Deception

- Caminada's notion of deception relies on the nonmonotonic inference capabilities of the hearer to reach a wrong conclusion.
- In [Sakama, Caminada, Herzig, JELIA 2010], it is formulated as

$$\begin{aligned} \text{DEC}_{ab}(\sigma, \delta) = & \text{utter}_{ab}(\sigma) \wedge B_a \sigma \wedge I_a B_b \sigma \\ & \wedge B_a B_b ((\sigma \wedge \neg B_b \neg \delta) \supset \delta) \\ & \wedge B_a \neg B_b \neg \delta \wedge B_a \neg \delta \\ & \wedge \neg \text{utter}_{ab}(\neg \delta) \wedge I_a B_b \delta \end{aligned}$$

- By the definition,  $\text{DEC}_{ab}(\sigma, \delta)$  implies  $WI_{ab}(\delta)$ .
- It does not describe the effect of deception, and is distinguished as **attempted deception**.

# Intended Deception, Attempted Deception, Lying and Withholding Information

