Abduction
Conversational Implicature
and Misleading

Chiaki Sakama (Wakayama University)

Katsumi Inoue (National Institute of Informatics)

MBR 2015

Abduction in Dialogue (1)

Mary: You're late this morning, aren't you?

John: The trains are not running on schedule.

From the response by John, Mary thinks that there was some accident.

- In this reasoning, Mary uses the implication

  "some accident happens" $\supset$ "trains are not running".

- Mary believes John's utterance and she has no reason to believe the negation of "some accident happens".

- Then Mary abduces "some accident happens" for an explanation of "trains are not running".

Objective Abduction

- Let a be a hearer and b a speaker. When b utters a (propositional) sentence φ, a (propositional) sentence ψ is inferred by **objective abduction** (**O-abduction**) from φ by a if

$$B_a\, \varphi \;\wedge\; B_a\, (\psi \supset \varphi) \;\wedge\; \neg\, B_a \neg \psi$$

  where $B_a\, \varphi$ means a believes φ.

- ψ is called an **O-explanation** of φ.
  We write O-abd$_a$ (φ, ψ) if ψ is an O-explanation of φ by a.

- It is called ``objective'' abduction because abduction is performed based on the objective fact of an utterance.

Abduction in Dialogue (2)

Mother: What are you doing?

Daughter: I'm writing a letter to Santa Claus.

From the response by her daughter, mother thinks that her daughter **believes** the existence of Santa Claus.

- Mother believes that her daughter **believes** the implication

  "Santa Claus exists" $\supset$ "She can write a letter to him".

- Mother believes that her daughter **believes** that she can write a letter to Santa Claus.

- Mother has no reason to believe that her daughter **disbelieves** the existence of Santa Claus.

- Then mother abduces "her daughter **believes** the existence of Santa Claus".

Subjective Abduction

- Let a be a hearer and b a speaker. When b utters a (propositional) sentence φ, a (propositional) sentence $B_b\,\psi$ is inferred by **subjective abduction** (**S-abduction**) from φ by a if

$$B_a\,B_b\,\varphi \;\wedge\; B_a\,B_b\,(\psi \supset \varphi) \;\wedge\; \neg\,B_a\neg\,B_b\,\psi$$

- $B_b\,\psi$ is called an **S-explanation** of φ.
  We write S-abd$_{ab}\,(\varphi,\,\psi)$ if $B_b\,\psi$ is an S-explanation of φ by a.

- It is called ``subjective'' abduction because abduction is performed based on the hearer's subjective view on the speaker's belief state.

O-abduction vs. S-abduction

- a: a hearer, b: speaker, φ: utterance
  - O-abduction infers ψ if $B_a\,\varphi \wedge B_a\,(\psi \supset \varphi) \wedge \neg B_a \neg \psi$
  - S-abduction infers $B_b\,\psi$ if $B_a\,B_b\,\varphi \wedge B_a\,B_b\,(\psi \supset \varphi) \wedge \neg B_a \neg B_b\,\psi$

- In O-abduction, a hearer a may believe an O-explanation ψ which accounts for an utterance φ by a speaker.

- In S-abduction, a hearer a may believe an S-explanation $B_b\,\psi$ but does not necessarily believe ψ by himself/herself.

O-abduction vs. S-abduction

- Suppose a speaker b utters his/her belief $B_b\,\varphi$.
  Then a hearer a infers $B_b\psi$ by O-abd$_a$ $(B_b\,\varphi,\,B_b\,\psi)$

  $$B_a\,B_b\,\varphi \;\wedge\; B_a\,(B_b\,\psi \supset B_b\,\varphi) \;\wedge\; \neg\,B_a\neg B_b\,\psi$$

- It should be distinguished from S-abd$_{ab}$ $(\varphi,\,\psi)$ which infer $B_b\psi$ by
  $$B_a\,B_b\,\varphi \;\wedge\; B_a\,B_b\,(\psi \supset \varphi) \;\wedge\; \neg\,B_a\neg B_b\,\psi$$

- Since $B_b\,(\psi \supset \varphi)$ implies $(B_b\,\psi \supset B_b\,\varphi)$,
  $$\text{S-abd}_{ab}\,(\varphi,\,\psi) \ \text{ implies } \ \text{O-abd}_a\,(B_b\,\varphi,\,B_b\,\psi)$$

- We can also consider another abduction O-abd$_a$ $(\varphi,\,B_b\psi)$ as
  $$B_a\,\varphi \;\wedge\; B_a\,(B_b\,\psi \supset \varphi) \;\wedge\; \neg\,B_a\neg B_b\,\psi$$

Different Types of Abduction in Dialogue

a: hearer
b: speaker

imply

O-abduction

$$\frac{\varphi \qquad \psi \supset \varphi}{\psi}$$

φ : utterance

$\psi \supset \varphi$: hearer a's belief

ψ : hearer a's explanation

S-abduction

$$\frac{\varphi \qquad B_b\,(\psi \supset \varphi)}{B_b\,\psi}$$

φ : utterance

$B_b\,(\psi \supset \varphi)$: hearer a's belief

$B_b\,\psi$: hearer a's explanation

O-abduction

$$\frac{B_b\,\varphi \qquad B_b\,\psi \supset B_b\,\varphi}{B_b\,\psi}$$

$B_b\,\varphi$: utterance

$B_b\,\psi \supset B_b\,\varphi$: hearer a's belief

$B_b\,\psi$: hearer a's explanation

O-abduction

$$\frac{\varphi \qquad B_b\,\psi \supset \varphi}{B_b\,\psi}$$

φ : utterance

$B_b\,\psi \supset \varphi$: hearer a's belief

$B_b\,\psi$: hearer a's explanation

Conversational Implicature (CI)

- A pragmatic inference to an implicit meaning of a sentence that is not actually uttered by a speaker (Grice 1975).

- Two principles from the speaker's viewpoints:
  - **Q-principle**: Say as much as you can.
  - **I-principle**: Say no more than you must.

- Two principles from the hearer's viewpoints:
  - **Q-implicature**: Imply the negation of a semantically stronger sentence than what is actually uttered.
  - **I-implicature**: Imply a semantically stronger (or more specific) sentence than what is actually uttered.

- Two implicatures conflict in their interpretations.

Examples

- ``I have two children" Q-implicates
  ``I do **not** have more than two children".

- ``Some of my friends like classical music" Q-implicates
  ``**Not all** of my friends like classical music".

- ``I will study French or Germany in the class" Q-implicates
  ``I will **not** study **both** French **and** Germany in the class".

- ``I have two dollars to pay the bill" I-implicates
  ``I have **at least** two dollars to pay the bill."

- ``We'll go on a picnic if it is fine tomorrow" I-implicates
  ``We'll go on a picnic if **and only if** it is fine tomorrow."

- ``John came to the office and he turns on the PC" I-implicates
  ``John came to the office and he, **John**, turns on the PC."

Conflicts between Abduction, Q-implicature and I-implicature

a: hearer
b: speaker

O-abduction

$$\frac{\varphi \qquad \psi \supset \varphi}{\psi}$$

φ : utterance

$\psi \supset \varphi$: hearer a's belief

ψ : hearer a's explanation

S-abduction

$$\frac{\varphi \qquad B_b\,(\psi \supset \varphi)}{B_b\,\psi}$$

φ : utterance

$B_b\,(\psi \supset \varphi)$: hearer a's belief

$B_b\,\psi$: hearer a's explanation

Q-implicature

$$\frac{\varphi \qquad B_b\,(\psi \supset \varphi)}{B_b\,\neg\psi}$$

φ : utterance

$B_b\,(\psi \supset \varphi)$: hearer a's belief

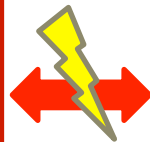$B_b\,\neg\psi$: hearer a's interpretation

I-implicature

$$\frac{\varphi \qquad B_b\,(\psi \supset \varphi)}{B_b\,\psi}$$

φ : utterance

$B_b\,(\psi \supset \varphi)$: hearer a's belief

$B_b\,\psi$: hearer a's interpretation

Q- and I-implicatures

- Let a be a hearer and b a speaker. When b utters a (propositional) sentence φ, a (propositional) sentence $B_b \neg \psi$ is inferred by **Q-implicature** from φ by a if
$$B_a B_b \varphi \wedge C(\psi \supset \varphi) \wedge \neg B_a B_b \psi$$

  On the other hand, a (propositional) sentence $B_b \psi$ is inferred by **I-implicature** from φ by a if
$$B_a B_b \varphi \wedge C(\psi \supset \varphi) \wedge \neg B_a \neg B_b \psi$$

- $C(\psi \supset \varphi)$ means that $\psi \supset \varphi$ is **common knowledge** that is shared by the speaker and the hearer.

- We write Q-imp$_{ab}(\varphi, \psi)$ (resp. I-imp$_{ab}(\varphi, \psi)$) if $B_b \neg \psi$ (resp. $B_b \psi$) is inferred by Q- (resp. I-) implicature from φ by a.

Why common knowledge in CI?

- Conversational implicature is based on **common knowledge**, i.e., both a speaker and a hearer know the truth of the implication $\psi \supset \varphi$ and each one also knows that the other party knows the truth of the sentence.

- If the hearer does not know whether or not the speaker knows the implication, then the hearer cannot infer the intended meaning of the speaker's utterance.

- If the speaker does not know whether or not the hearer knows the implication, then the speaker cannot expect the hearer's reasoning by Q/I-implicature.

- Thus conversational implicature is in effect if and only if a speaker and a hearer share the same knowledge and each one knows that the other party also shares the same knowledge.

Abduction vs. CI (1)

- Both abduction and CI use an implication $\psi \supset \varphi$ to infer information behind an utterance. In abduction, the implication is a hearer's **private belief**, while in CI it is **common knowledge**.

- Abduction is a process of private reasoning, and one can reason abductively without knowing the belief state of the other party. By contrast, conversation aims at communicating information.

- Since "$C \varphi \supset B_a \varphi$" and "$C \varphi \supset B_a B_b \varphi$", one may use common knowledge for the purpose of abduction, but not vice versa.

Abduction vs. CI (2)

- a: a hearer, b: speaker, φ: utterance

  I-imp$_{ab}$ (φ, ψ): infer $B_b\psi$ from $B_a\,B_b\,\varphi \wedge C\,(\psi \supset \varphi) \wedge \neg\,B_a\neg B_b\,\psi$
  S-abd$_{ab}$ (φ, ψ): infer $B_b\psi$ from $B_a\,B_b\,\varphi \wedge B_a\,B_b\,(\psi \supset \varphi) \wedge \neg\,B_a\neg B_b\,\psi$

- Since $C\,(\psi \supset \varphi)$ implies $B_a\,B_b\,(\psi \supset \varphi)$,

  I-imp$_{ab}$ (φ, ψ) implies S-abd$_{ab}$ (φ, ψ).

- Q-imp$_{ab}$ (φ, ψ) : infer $B_b\neg\psi$ from $B_a\,B_b\,\varphi \wedge C\,(\psi \supset \varphi) \wedge \neg\,B_a\,B_b\,\psi$
  S-abd$_{ab}$ (φ, ψ): infer $B_b\psi$ from $B_a\,B_b\,\varphi \wedge B_a\,B_b\,(\psi \supset \varphi) \wedge \neg\,B_a\neg B_b\,\psi$

- When $B_a\,B_b\,\varphi \wedge C\,(\psi \supset \varphi)$,
  a hearer may conclude $B_b\,\psi$ by S-abduction if $\neg B_a\neg B_b\,\psi$; while
  a hearer may conclude $B_b\neg\psi$ by Q-implicature if $\neg B_a\,B_b\,\psi$.

Conflicts between Abduction, Q-implicature and I-implicature

O-abduction

$$\frac{\varphi \qquad \psi \supset \varphi}{\psi}$$

φ : utterance

$\psi \supset \varphi$: hearer a's belief

ψ : hearer a's explanation

a: **hearer**
b: **speaker**

S-abduction

$$\frac{\varphi \qquad B_b\,(\psi \supset \varphi)}{B_b\,\psi}$$

φ : utterance

$B_b\,(\psi \supset \varphi)$: hearer a's belief

$B_b\,\psi$: hearer a's explanation

imply

Q-implicature

$$\frac{\varphi \qquad C(\psi \supset \varphi)}{B_b \neg \psi}$$

φ : utterance

$C(\psi \supset \varphi)$: common knowledge

$B_b \neg \psi$: hearer a's interpretation

I-implicature

$$\frac{\varphi \qquad C(\psi \supset \varphi)}{B_b\,\psi}$$

φ : utterance

$C(\psi \supset \varphi)$: common knowledge

$B_b\,\psi$: hearer a's interpretation

What happens if a hearer does not believe an utterance?

- Suppose the Turing's imitation game in which a human judge asks questions to an interlocutor in order to determine whether he or she is interacting with a human or a machine.

  Judge (*a*): Are you a machine?
  Interlocutor (*b*): I'm a human.

- Suppose that the judge believes the implication:
  $$\neg\text{machine} \supset \text{human}$$
  (The interlocutor is human if he/she is not a machine.)

- Given the response ``human" by the interlocutor, will the judge believe that the interlocutor is not a machine (by O-abduction)?

What happens if a hearer does not believe an utterance?

- In the Turing imitation game, a machine attempts to convince a judge that it is human through appropriate, and often **deceptive** responses.

- In the above dialogue, if the judge disbelieves the utterance φ by the interlocutor, then $\neg \, B_a \, \varphi$ holds thereby O-abd$_a$ (φ, ψ)

$$B_a \, \varphi \, \wedge \, B_a \, (\psi \supset \, \varphi) \, \wedge \neg \, B_a \neg \psi$$

where φ=human, ψ= $\neg$ machine

is not applied and the judge does not abduce ψ= $\neg$ *machine.*

What happens if a hearer believes the falsity of an utterance?

- Suppose the same dialogue

  Judge (a): Are you a machine?

  Interlocutor (b): I'm a human.

  and the judge believes $\neg$machine $\supset$ human as before.

- The interlocutor utters φ =*human*, but the judge believes the contrary $\neg\varphi$.

- In this case, it holds that $B_a \neg \varphi \wedge B_a (\psi \supset \varphi) \supset B_a \neg \psi$ and the judge believes $\neg \psi$=*machine*.

What happens if a hearer believes that a speaker is lying?

- Dialogue:

  Judge (a): Are you a machine?
  Interlocutor (b): I'm a human.

- The judge believes that the interlocutor believes the implication $\psi \supset \varphi = (\neg machine \supset human)$ and the judge also believes that the interlocutor is **lying**, i.e., the judge believes that the interlocutor believes the falsity of his/her utterance $\varphi = human$.

- In this case, it holds that
  $$B_a B_b \neg\varphi \wedge B_a B_b (\psi \supset \varphi) \supset B_a B_b \neg\psi$$
  then the judge a believes that the interlocutor believes $\neg\psi = machine$.

Misleading

- A speaker may believe that a hearer would abduce ψ as a result of the speaker's utterance φ.
- Consider the dialogue.

  Judge (a): Are you a machine?

  Interlocutor (b): Shall I sing a song?

- The interlocutor (who is in fact a human) expects that his/her response would make the judge abduce the fact ``human'' based on his/her belief that the judge believes the implication ``human $\supset$ sing''.
- Thus a speaker will decide what to say by considering the effect of his/her utterance on the hearer's side.
- A speaker may use this to mislead a hearer to reach a wrong assumption.

Misleading by O-abduction

- When a speaker b utters a sentence φ to a hearer a,
  b **misleads** a **by O-abduction** if
  $$B_b\,(B_a\,\varphi \wedge B_a\,(\psi \supset \varphi) \wedge \neg\, B_a \neg\psi) \wedge B_b \neg\psi$$

- We write O-mislead$_{ba}$ $(\varphi,\, \psi)$ if b's utterance φ misleads a to abduce an O-explanation ψ .

- The above formula says that a speaker b believes that his/her utterance would lead a hearer a to an assumption ψ by O-abduction, however, b believes $\neg\psi$.

- A speaker may use a weaker version of misleading by replacing $B_b \neg\psi$ with $\neg B_b \psi$.

Misleading by O-abduction

- Dialogue:
  Judge (a): Are you a machine?
  Interlocutor (b): I'm a human.

- The interlocutor (who is in fact a machine) believes that the judge a believes the response φ =human by b .

- The interlocutor also believes that:
  the judge believes the implication $\neg$machine $\supset$ human
  while disbelieves $\neg\psi$=machine.

- If the interlocutor believes that it is a machine $\neg\psi$=machine, the interlocutor misleads the judge by the response φ =human.

- A speaker's utterance will change depending on his/her belief that whether a hearer believes the speaker's utterance or not.

- Suppose that the interlocutor is a machine and it considers that the judge will doubt its response. In this situation, consider the dialogue
  Judge (a): Are you a machine
  Interlocutor (b): Yes, I'm a machine.

- If the judge believes the falsity of the utterance, he/she interprets the contrary of the response and concludes the interlocutor is a human.
  $B_a \neg machine \wedge B_a(\neg machine \supset human) \supset B_a human$

- However, this is what the interlocutor has intended. In this case, the interlocutor reasons by the formula:

  $B_b B_a \neg machine \wedge B_b B_a (\neg human \supset machine) \supset B_b B_a human.$

- The interlocutor b believes that the judge a believes the contrary of the utterance *machine*, expecting that the judge reaches the wrong conclusion *human* using the implication $\neg human \supset machine$.

Misleading by S-abduction

- When a speaker b utters a sentence φ to a hearer a,
  b **misleads** a **by S-abduction** if
  $$B_b\,(B_a\,B_b\,\varphi\ \wedge\ B_a\,B_b(\psi\supset\varphi)\ \wedge\neg\,B_a\neg\ B_b\,\psi)\ \wedge\ B_b\neg B_b\,\psi$$

- We write S-mislead$_{ba}\,(\varphi,\ \psi)$ if b's utterance φ misleads a to abduce an S-explanation $B_b\,\psi$.

- The above formula says that a speaker b believes that his/her utterance would lead a hearer a to an assumption $B_b\,\psi$ by S-abduction, however, b believes $\neg\,B_b\,\psi$.

- S-mislead$_{ba}(\varphi,\ \psi)$ implies O-mislead$_{ba}\,(B_b\,\varphi,\ B_b\,\psi)$.

Misleading by telling the truth

- A speaker may utter what he/she believes true while expecting a hearer will make an incorrect abduction.

- When an interlocutor is a machine, suppose the dialogue
  Judge (a): Are you a machine?
  Interlocutor (b): I often make errors.

- The interlocutor expects that the judge will consider it a human by the implication $human \supset error$.

- However, the interlocutor (machine) in fact often makes calculation errors by programming bugs, so it tells the truth.

- Such a speech act is often said ``**indirect lies**'' or ``**lying while saying the truth**''.

Misleading by Q-implicature

- Dialogue:

  Mother (a): How was your math exam?
  Son (b): I could not solve one question.

- Using Q-implicature, mother believes that her son worked out other questions, except the one that could not be solved.

- However, this is what the son has intended. In fact, he believes that he could not solve more than one question.

- Since he believes that he could not solve more than one question, he also believes that he could not solve one question. He then uttered his weaker belief in response to her question.

- On the other hand, mother believes that his utterance must mean that he does not have failed more than one question.

Misleading by CI

- When a speaker b utters a sentence φ to a hearer a,
  b **misleads** a **by Q-implicature** if
  $$B_b\,(B_a\,B_b\,\varphi \;\wedge\; C\,(\psi \supset \varphi) \;\wedge\; \neg B_a\,B_b\,\psi) \;\wedge\; B_b\psi$$

- A speaker b believes that the utterance φ leads a hearer a to conclude the negation of a stronger sentence $\neg\psi$ by Q-implicature, while b believes ψ .

- Likewise, b **misleads** a **by I-implicature** if
  $$B_b\,(B_a\,B_b\,\varphi \;\wedge\; C\,(\psi \supset \varphi) \;\wedge\; \neg B_a\,\neg B_b\,\psi) \;\wedge\; \neg B_b\psi$$

- A speaker b believes that the utterance φ leads a hearer a to conclude a weaker sentence ψ by I-implicature, while b disbelieves ψ .

Abduction vs. CI in Misleading

- Misleading by conversational implicature may fail if a speaker believes that a hearer uses Q-implicature (resp. I-implicature) but in fact the hearer uses I-implicature (resp. Q-implicature).

- When a speaker b disbelieves a sentence ψ, he/she would have two options for misleading a hearer a to believe $B_b\psi$.

- One is uttering φ under the condition that there is common knowledge $C\,(\psi \supset \varphi)$ and the speaker believes that the hearer uses I-implicature (i.e., misleading by I-implicature).

- The other is uttering φ under the condition that the speaker believes $B_b\,(\psi \supset \varphi)$ and that the speaker believes that the hearer uses S-abduction. (i.e., misleading by S-abduction).

Final Remark

- Two different types of abduction and two different conversational implicatures (CIs) are formulated using propositional modal logic.

- Abduction uses private belief of a reasoner, while CI relies on common knowledge between participants in a conversation.

- We also argued how a speaker would mislead a hearer in conversation.

- The framework is simple but capable of capturing different aspects of abduction and CI in human dialogues, that have not been thoroughly investigated in the literature.