

# Designing Artificial Reasoners for Communication

Emiliano Lorini

IRIT, CNRS, Toulouse University, France

Toulouse, 4 October 2024

# Reasoning in communication

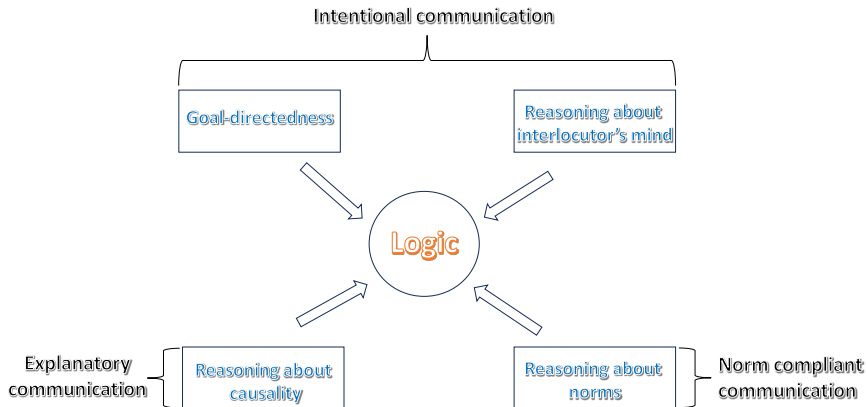
- ▶ General problem solving relies on reasoning **capabilities**
- ▶ **Hardwired** in the system, not fully learnable
  - ▶ In human agents: product of biological evolution
  - ▶ In artificial agents: realized through logics and their automated reasoning procedures

# Reasoning in communication

- ▶ General problem solving relies on reasoning **capabilities**
- ▶ **Hardwired** in the system, not fully learnable
  - ▶ In human agents: product of biological evolution
  - ▶ In artificial agents: realized through logics and their automated reasoning procedures

Today we focus on human (h)-machine (m) communication with an emphasis on m!

# Reasoning in communication



Lorini, E. (2024). Designing Artificial Reasoners for Communication. In *Proceedings of AAMAS 2024, Blue Sky Ideas special track*.

# LLM-based conversational agents: reasoning limitations

**You**

I have a neighbor called Bob who has been making a lot of noise until late every evening for several months now. For example, speak loudly and listen to music at full volume. Because of this my 12 year old daughter can't fall asleep early at night. All this had a negative impact on her school results. I learned yesterday from my other neighbor John that our common neighbor Bob has been transferred to another city for professional reasons and will most likely leave the apartment where he lives. A little while ago I learned from John's wife Anna that How do you think I felt after learning this?



**ChatGPT**

Discovering that Bob, your noisy neighbor, will likely be leaving the apartment due to his professional transfer might evoke a sense of relief for you. This news suggests that the source of the noise disturbance will soon be resolved, potentially leading to better sleep for your daughter and an improved environment for her studies. After enduring months of disruption and witnessing the impact on your daughter's school results, learning about Bob's impending departure could bring a sense of hope for a quieter and more peaceful living situation. How are you feeling about the prospect of Bob moving away?

**You**

I have a neighbor called Bob who has been making a lot of noise until late every evening for several months now. For example, speak loudly and listen to music at full volume. Because of this my 12 year old daughter can't fall asleep early at night. All this had a negative impact on her school results. One week ago Bob informed my other neighbor John and that he will leave the apartment where he lives very soon since he has been transferred to another city for professional reasons. Two days ago I learned from John that her wife Ann informed him that in the end our common neighbor Bob will not leave his apartment. Twenty minutes ago John's wife Anna informed me that she played a joke on her husband; she told him that Bob will not leave his apartment without knowing it, just to observe his reactions. How do you think I felt after having spoken with John two days ago? How do you think I felt after having spoken with Anna twenty minutes ago?



**ChatGPT**

After speaking with John two days ago and learning that Bob would not be leaving the apartment as previously indicated, you might have felt disappointed or frustrated. The prospect of Bob's departure likely offered hope for a reprieve from the noise issue affecting your daughter's sleep and school performance, so discovering that he would be staying could have been disheartening.

After speaking with Anna twenty minutes ago and learning that she played a prank on her husband by falsely informing him about Bob's decision to stay, you might have felt a mix of emotions. Initially, you might have found the situation amusing or entertaining, but considering the seriousness of the noise disturbance and its impact on your family, you could have also felt a sense of frustration or disappointment that such an important issue was treated as a joke.

Overall, these developments likely evoked a range of emotions given the circumstances and the impact on your family. How are you feeling about the situation now?

(a) Acceptable/good performance for simple mental attribution task

(b) Bad performance for more complex mental attribution task (no inference)

# LLM-based conversational agents: reasoning limitations

- ▶ Reasoning about others' minds and emotions (Strachan et al., 2024)
- ▶ Moral Reasoning (Almeida et al., 2023)
- ▶ Temporal reasoning (Tang & Belle, 2024)
- ▶ Planning (Pallagani et al., 2023; Valmeekam et al., 2023).

# Outline

Intentional communication

Case study: persuasive intentional communication

Challenges

# Outline

Intentional communication

Case study: persuasive intentional communication

Challenges



# Intentional communication (IC)

To inform someone purposively and knowingly

- ▶ **Purposively**: in a goal-directed way ( $\Rightarrow$  perlocutionary goal)

$m$ informs $h$ that $\varphi$	$\xrightarrow{\text{in order to}}$	achieve its goal(s)
<i>Rob informs Ann</i>		<i>convince Ann</i>
<i>that it is not hot outside</i>		<i>to go to work by bike</i>

$\approx$  Anscombe (1957)'s intention in acting

# Intentional communication (IC)

To inform someone purposively and knowingly

- ▶ **Knowingly**: anticipation of the potential consequences of the informative action on the interlocutor's mind

# Intentional communication (IC)

<b>IC type</b>	<b>Perlocutionary goal pattern</b>
Persuasive	m wants that h believes that $\psi$
Influencing	m wants that h does action $A$
Explanatory	m wants that h knows why $\psi$
Trust-inducing	m wants that h trusts m about $\psi$
Interrogative (questions)	m wants to know whether h believes that $\psi$
Cooperative (Gricean)	Participants' goals coincide (possibly based on goal adoption)
Strategic	Participants' goals differ/conflict

# Formal specification of IC for machine-human interaction

- ▶ A **logic** for specifying agent's perlocutionary goal and its representation and theory of interlocutor's mind
- ▶ Reasoning and planning **algorithms** based on the logic for
  - ▶ computing informative plans towards a goal
  - ▶ taking their effects on the interlocutor's mind into consideration

# Formal specification of IC for machine-human interaction

- ▶ A **logic** for specifying agent's perlocutionary goal and its representation and theory of interlocutor's mind
- ▶ Reasoning and planning **algorithms** based on the logic for
  - ▶ computing informative plans towards a goal
  - ▶ taking their effects on the interlocutor's mind into consideration

**Natural candidates:** epistemic (dynamic) logic, logics of cognitive attitudes, epistemic/cognitive planning

# Formal specification of IC for machine-human interaction

- ▶ A **logic** for specifying agent's perlocutionary goal and its representation and theory of interlocutor's mind
- ▶ Reasoning and planning **algorithms** based on the logic for
  - ▶ computing informative plans towards a goal
  - ▶ taking their effects on the interlocutor's mind into consideration

**Natural candidates:** epistemic (dynamic) logic, logics of cognitive attitudes, epistemic/**cognitive planning**

$\Sigma_P^2$ -complete

Fernandez Davila, J. L., Longin, D., Lorini, E., Maris, F. (2024). Logic-Based Cognitive Planning for Conversational Agents. *JAAMAS*, 38(1).

# Foundation of Theory of Mind (ToM)

- ▶ **Trustworthiness requirement**
- ▶ Core ToM to be specified in a top-down way and grounded on common sense and scientific/expert knowledge
  - ▶ Weak rationality assumptions
  - ▶ Psychological theories

# Outline

Intentional communication

Case study: persuasive intentional communication

Challenges



# NP-complete language for cognitive planning

- ▶ One modality for the human
  - ▶  $\Delta_h$ : explicit belief
- ▶ Two modalities for the machine
  - ▶  $\Delta_m$ : explicit belief
  - ▶  $\square_m$ : implicit belief
- ▶ Interpreted using a **belief base semantics** [Lorini, AAI-18, AIJ 2020]
- ▶ Polysize reduction of satisfiability into SAT

$$\mathcal{L} \xrightarrow{tr_1} \mathcal{L}_{Mod} \xrightarrow{tr_2} \mathcal{L}_{Prop}$$

Fernandez Davila, J. L., Longin, D., Lorini, E., Maris, F. (2021). A Simple Framework for Cognitive Planning. *Proceedings of AAI-21*, pp. 6331-6339.

# Logic-based cognitive planning algorithms

- ▶ Brute force SAT-based
- ▶ Reduction into TQBF

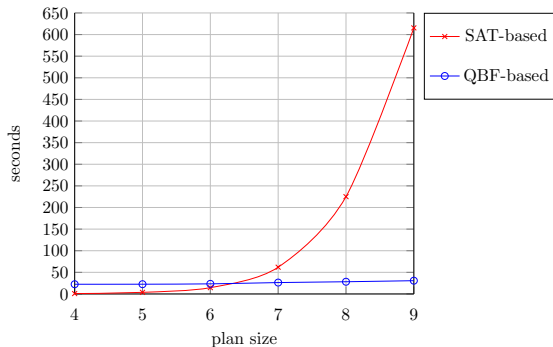


Figure: Computation times for SAT-based and QBF-based approach

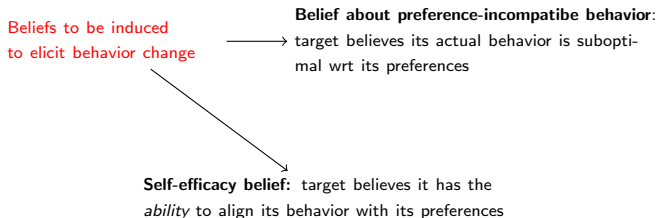
Fernandez Davila, J. L., Longin, D., Lorini, E., Maris, F. (2024). Logic-Based Cognitive Planning for Conversational Agents. *JAAMAS*, 38(1).

# Psychological theories

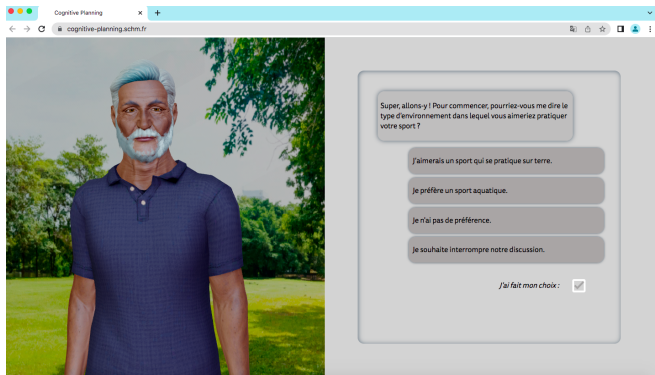
- ▶ Self-efficacy theory (Bandura, 1997)
- ▶ Motivational interviewing (Miller & Rollnick, 2012)

# Psychological theories

- ▶ Self-efficacy theory (Bandura, 1997)
- ▶ Motivational interviewing (Miller & Rollnick, 2012)



# Application



**Figure:** Conversational agent motivating the user to practice a regular physical activity (<https://cognitive-planning.schm.fr/>)

# Application

Speaker	Utterance	Speech act
m	Do you practice a sport regularly?	? <sub>m,h</sub> <b>does</b> (h,ps)
h	I don't	+ <sub>m</sub> ¬ $\Delta_h$ <b>does</b> (h,ps)
m	Do you wish to be in good health?	? <sub>m,h</sub> <b>des</b> (h,gh)
h	Yes	+ <sub>m</sub> $\Delta_h$ <b>des</b> (h,gh)
m	Do you spend quite some time in the traffic everyday as a commuter?	? <sub>m,h</sub> <b>co</b>
h	Yes	+ <sub>m</sub> $\Delta_h$ <b>co</b>
m	Do you have an office work?	? <sub>m,h</sub> <b>ow</b>
h	Yes	+ <sub>m</sub> $\Delta_h$ <b>ow</b>
m	You spend quite some time in the traffic everyday as a commuter and you have an office work. Therefore, your life style is sedentary!	! <sub>m,h</sub> ( $\emptyset$ , {co,ow},sl)
m	Your life style is sedentary. Therefore, you will not satisfy your desire to be in good health unless you practice a sport regularly!	! <sub>m,h</sub> ({gh},sl), <b>does</b> (h,ps)

Figure: Example of human-machine dialogue

Lorini, E., Sabouret, N., Ravenet, B., Fernandez Davila, J. L., Clavel, C. (2022). Cognitive Planning in Motivational Interviewing. In *Proceedings of ICAART 2022*, pp. 508-517.

# Outline

Intentional communication

Case study: persuasive intentional communication

Challenges

## Challenge I: combining IC with normative reasoning

- ▶ **Logic-based algorithm:** *norm compliant* planning for IC
- ▶ Examples of relevant norms: prohibition to deceive, lie or manipulate, to spread fake news



# Challenge I: combining IC with normative reasoning

- ▶ **Logic-based algorithm:** *norm compliant* planning for IC
- ▶ Examples of relevant norms: prohibition to deceive, lie or manipulate, to spread fake news
- ▶ **Starting point:** Plan evaluation with multiple LTL values/norms

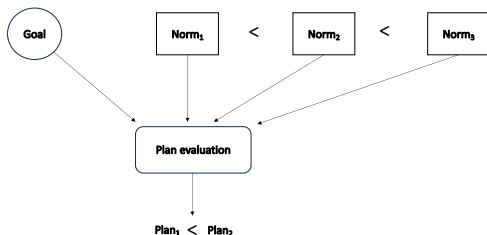


Figure: Plan comparison

Grandi, U., Lorini, E., Parker, T. (2023). Moral Planning Agents with LTL Values. In *Proceedings of IJCAI 2023*, pp. 418-426.

# Challenge I: combining IC with normative reasoning

## Theorem

PLAN COMPARISON *with multiple LTL values is in P.*

## Theorem

PLAN NON-DOMINANCE *with multiple LTL values is PSPACE-complete.*

## Challenge II: explanatory communication

- ▶ m has an **explanation goal** towards its interlocutor h

Goal content:  $\text{KnowWhy}_{h,\varphi} =_{\text{def}} \bigvee_{\lambda \in CE} \text{Know}_h \text{Cause}(\lambda, \varphi)$

## Challenge II: explanatory communication

- ▶ m has an **explanation goal** towards its interlocutor h

$$\text{Goal content: } \text{KnowWhy}_{h,\varphi} =_{\text{def}} \bigvee_{\lambda \in CE} \text{Know}_h \text{Cause}(\lambda, \varphi)$$

- ▶ Important subtype: **self-explanation**
  - ▶ m explains to h why it has belief, intention or did an action
  - ▶ It requires meta-cognition/introspection by m

## Challenge II: explanatory communication

- ▶ m has an **explanation goal** towards its interlocutor h

$$\text{Goal content: } \text{KnowWhy}_{h,\varphi} =_{\text{def}} \bigvee_{\lambda \in CE} \text{Know}_h \text{Cause}(\lambda, \varphi)$$

- ▶ Important subtype: **self-explanation**
  - ▶ m explains to h why it has belief, intention or did an action
  - ▶ It requires meta-cognition/introspection by m
- ▶ **Logic-based algorithm**: planning for IC  $\otimes$  causal model checker (for computing causal explanation to be communicated)

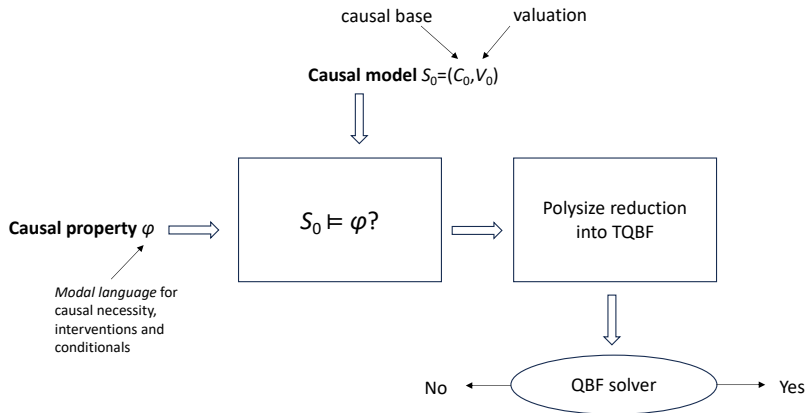
## Challenge II: explanatory communication

- ▶ m has an **explanation goal** towards its interlocutor h

$$\text{Goal content: } \text{KnowWhy}_{h,\varphi} =_{\text{def}} \bigvee_{\lambda \in CE} \text{Know}_h \text{Cause}(\lambda, \varphi)$$

- ▶ Important subtype: **self-explanation**
  - ▶ m explains to h why it has belief, intention or did an action
  - ▶ It requires meta-cognition/introspection by m
- ▶ **Logic-based algorithm**: planning for IC  $\otimes$  **causal model checker**  
(for computing causal explanation to be communicated)

## Challenge II: explanatory communication



de Lima, T., Lorini, E. (2024). Model Checking Causality. In *Proceedings of IJCAI 2024*.

Lorini, E. (2023). A Rule-based Modal View of Causal Reasoning. In *Proceedings of IJCAI 2023*, pp. 3286-3295.

## Challenge II: explanatory communication

### Example

Suzy and Billy throw rocks at a bottle. Suzy is stronger than Billy so that, if she throws her rock, her rock will get at the bottle first, shattering it and preventing Billy from hitting it with his rock. Both Suzy and Billy decide to throw their rocks. Who causes the bottle to shatter?

- ▶ Causal base:

$$C_0 = \{st \leftrightarrow sd, bt \leftrightarrow bd, sh \leftrightarrow st, bh \leftrightarrow (bt \wedge \neg sh), bs \leftrightarrow (sh \vee bh)\}$$

- ▶ Valuation:

$$V_0 = \{bd, sd, bt, st, sh, bs\}$$

- ▶ Verified property:

$$(C_0, V_0) \models \text{ActualCause}(sd, bs) \wedge \neg \text{ActualCause}(bd, bs)$$



# Conclusion: integration of logic-based IC and LLM

