Observable and Attention-Directing BDI Agents for Human-Autonomy Teaming

Blair Archibald, Muffy Calder, Michele Sevegnani, **Mengwei Xu**

Project: Multi-Perspective Design of IoT Cybersecurity in Ground and Aerial Vehicles (funded by Petras)
Project: Science of Sensor System Software (funded by EPSRC)

University of Glasgow

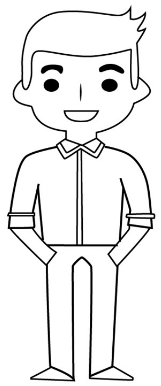
Human Autonomy Teaming

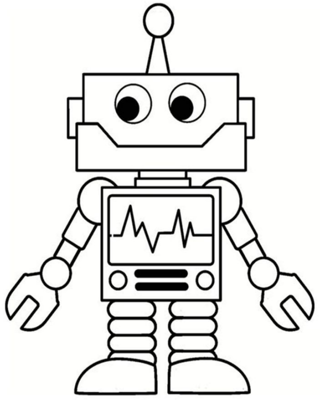
Definition:

A term describing

humans,
autonomous agents,

working together to achieve some objectives





Thomas O'Neill et al. (2020): Human-Autonomy Teaming: A Review and Analysis of Empirical Literature. Human Factors

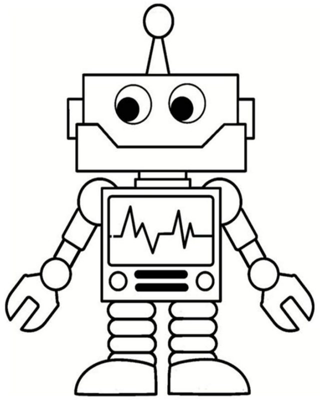
Human Autonomy Teaming

Building suitable agents for human-autonomy teaming :

A term describing

humans,
autonomous agents,

working together to achieve some objectives

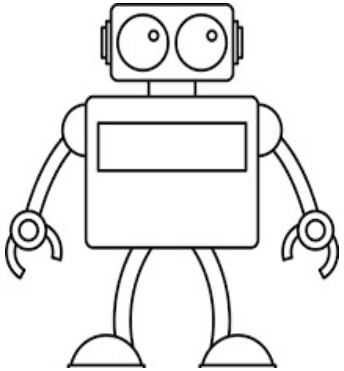


Thomas O'Neill et al. (2020): Human-Autonomy Teaming: A Review and Analysis of Empirical Literature. Human Factors

Building Suitable Agents for Human-Autonomy Teaming

Question:

what do humans want the agent to tell them as it is working for or with them?

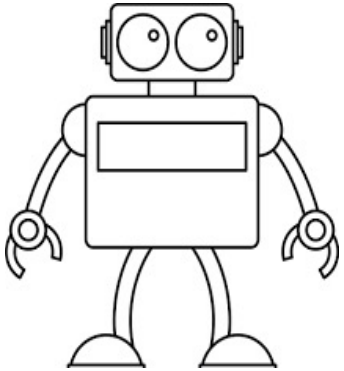


Building Suitable Agents for Human-Autonomy Teaming

Motivation:

what do humans want the agent to tell them as it is working for or with them?

so that



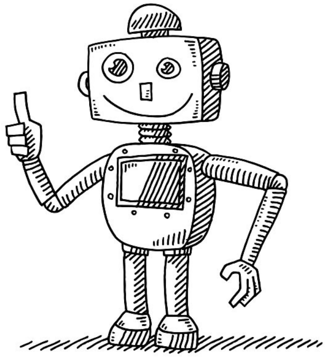
humans can partner effectively with the autonomy and understand what it is doing.

Building Suitable Agents for Human-Autonomy Teaming

Motivation:

humans can partner effectively with the autonomy and understand what it is doing?

entails



a shared understanding of the problem to be solved and progress toward goals

Building Suitable Agents for Human-Autonomy Teaming

Summary:

Question: what do humans want the agent to tell them as it is working for or with them?

Answers:

- Observability:
  - providing information of what an autonomy is doing relative to task progress

- Directing Attention:
  - directing the attention of the human to critical problems and changes.

Building Suitable Agents for Human-Autonomy Teaming

Summary:

Question: what do humans want the agent to tell them as it is working for or with them?

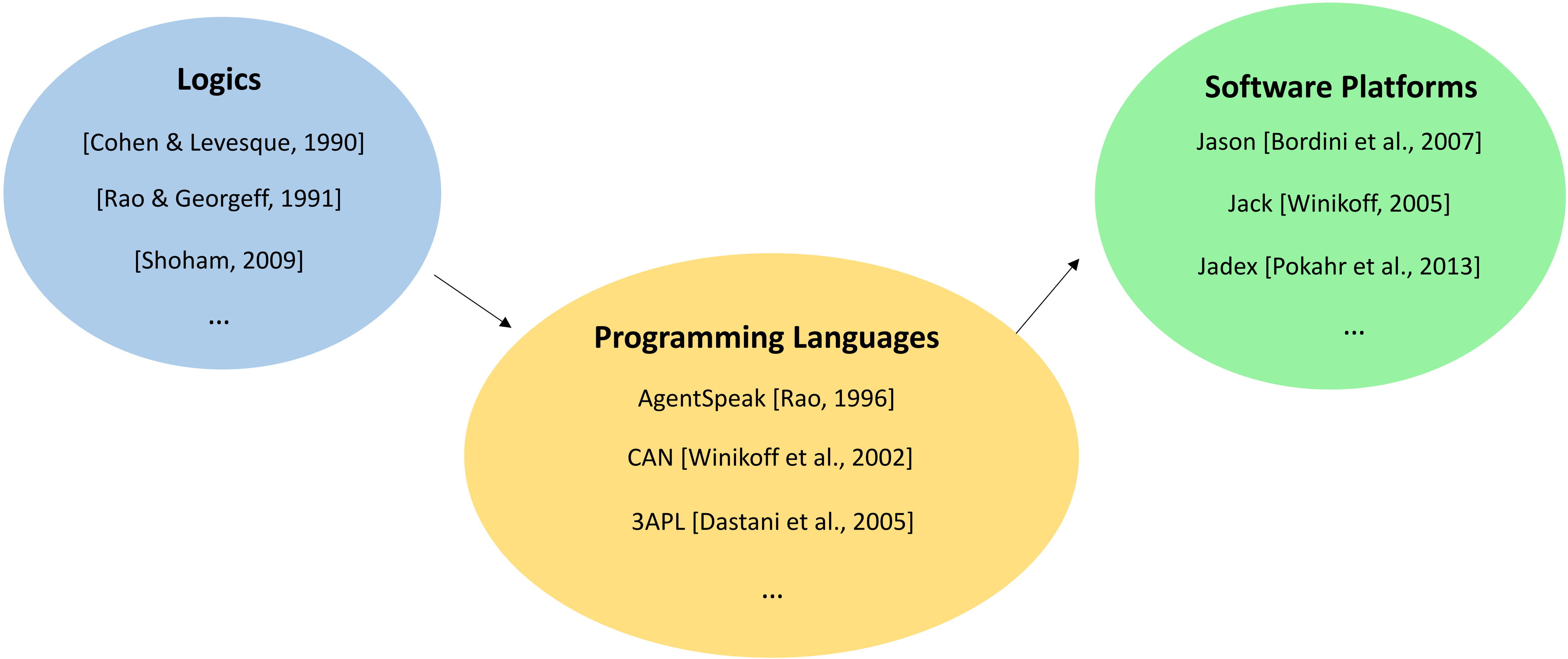
Answers:

- Observability:
  - providing information of what an autonomy is doing relative to task progress
- Directing Attention:
  - directing the attention of the human to critical problems and changes.

of course, this answer is the strict sub-set of the perfect answer; limitation and future of this work will be discussed in details later on

Building Suitable Agents for Human-Autonomy Teaming

Belief-Desire-Intention Framework:



Building Suitable Agents for Human-Autonomy Teaming

Belief-Desire-Intention Framework:

Programming Languages

AgentSpeak [Rao, 1996]

CAN [Winikoff et al., 2002]

CANPLAN [Sardina et al., 2011]

Conceptual Agent Notation (CAN)

Extension of Classical BDI Language: AgentSpeak

1. which is high-level programming that captures the essence of BDI concepts without implementation details, e.g. data structures
2. which provides formal and succinct operation semantics
3. which provides advanced behaviours including declarative goal, concurrency within an intention, and failure recovery.

Building Suitable Agents for Human-Autonomy Teaming

Belief-Desire-Intention Framework:

Programming Languages

AgentSpeak [Rao, 1996]

CAN [Winikoff et al., 2002]

CANPLAN [Sardina et al., 2011]

Conceptual Agent Notation (CAN)

Extension of Classical BDI Language: AgentSpeak

1. which is high-level programming that captures the essence of BDI concepts without implementation details, e.g. data structures
2. which provides formal and succinct operation semantics
3. which provides advanced behaviours including declarative goal, concurrency within an intention, and failure recovery.

Importantly, though we focus on CAN, the language features are similar to those of other mainstream BDI languages and the same modelling techniques would apply

Building Suitable Agents for Human-Autonomy Teaming

Belief-Desire-Intention Framework:

Recall

Questions: what do humans want the agent to tell them as it is working for or with them?

- Observability: providing information of what an autonomy is doing relative to task progress
- Directing Attention: directing the attention of the human to critical problems and changes.

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Questions: what do humans want the agent to tell them as it is working for or with them?

- Observability: providing information of what an autonomy is doing relative to task progress
- Directing Attention: directing the attention of the human to critical problems and changes.

Answers in BDI context:

Observability	Directing Attention
• provide information of status of agent's intentions • provide information of agent's progress of intentions	• direct attention to relevant environmental changes

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of status (e.g. success/failure) of agent's intentions

External Event: $e \in E^e$

e.g. new goal

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of status (e.g. success/failure) of agent's intentions

External Event: $e \in E^e$

e.g. new goal

External Event: $\langle e, I, status \rangle \in E^e$

the unique identifier

$status \in \{pending, active, success, fail\}$

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of status (e.g. success/failure) of agent's intentions

original CAN semantics

$$\frac{e \in E^e}{\langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e \backslash \{e\}, \mathcal{B}, \Gamma \cup \{e\} \rangle} \quad A_{event}$$

event selection

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of status (e.g. success/failure) of agent's intentions

original CAN semantics

$$\frac{e \in E^e}{\langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e \backslash \{e\}, \mathcal{B}, \Gamma \cup \{e\} \rangle} \quad A_{event}$$

event selection

new CAN semantics

$$\frac{\langle e, I, pending \rangle \in E^e}{\langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e \backslash \{\langle e, I, pending \rangle\} \cup \langle e, I, active \rangle, \mathcal{B}, \Gamma \cup \{\langle e, I \rangle\} \}} \quad A^{new}_{event}$$

1. switch an external event from pending to active,
2. link the intention with its original event with a same identifier

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of status (e.g. success/failure) of agent's intentions

original:

$$\frac{P \in \Gamma \quad \langle \mathcal{B}, P \rangle \nrightarrow}{\langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e, \mathcal{B}', \Gamma \backslash \{P\} \rangle} \quad A_{update}$$

Update external events

new:

$$\frac{\langle P, I \rangle \in \Gamma \quad \langle e, I, \textcolor{red}{active} \rangle \in E^e \quad \langle \mathcal{B}, \langle P, I \rangle \rangle \nrightarrow \quad P = nil}{\langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e \backslash \{\langle e, I, \textcolor{red}{active} \rangle\} \cup \langle e, I, \textcolor{red}{success} \rangle, \mathcal{B}, \Gamma \backslash \{\langle P, I \rangle\} \rangle} \quad A^{new}_{update\_suc}$$

$$\frac{\langle P, I \rangle \in \Gamma \quad \langle e, I, \textcolor{red}{active} \rangle \in E^e \quad \langle \mathcal{B}, \langle P, I \rangle \rangle \nrightarrow \quad P \neq nil}{\langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e \backslash \{\langle e, I, \textcolor{red}{active} \rangle\} \cup \langle e, I, ` \rangle, \mathcal{B}, \Gamma \backslash \{\langle P, I \rangle\} \rangle} \quad A^{new}_{update\_fail}$$

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of agent’s progress of intentions

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of agent's progress of intentions

intuition

plan $P_1 = e_1 : \varphi_1 \leftarrow a_1 ; a_2$

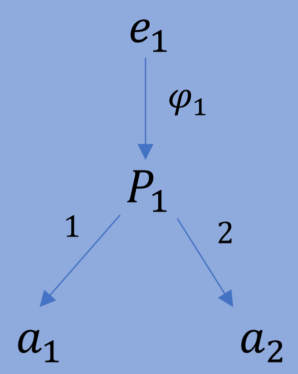
Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of agent's progress of intentions

intuition

plan $P_1 = e_1 : \varphi_1 \leftarrow a_1 ; a_2$



trace: e_1, P_1, a_1, a_2

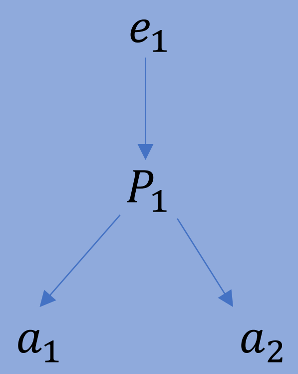
Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of agent's progress of intentions

intuition

plan $P_1 = e_1 \colon \varphi_1 \leftarrow a_1 ; a_2$



if the current step is at plan P_1
we can say that the progress is 2/4= 50%
where 2 is the position of P_1 and 4 is the length of the trace.

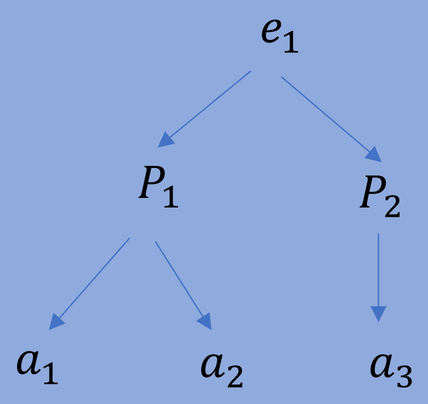
trace: e_1, P_1, a_1, a_2

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of agent's progress of intentions

multiple plans



if the current step is at event e_1
what do we say about the progress?
by average: 1/((3+4)/2)= 1/3.5= 28.5%

trace 1: e_1, P_1, a_1, a_2, length 4

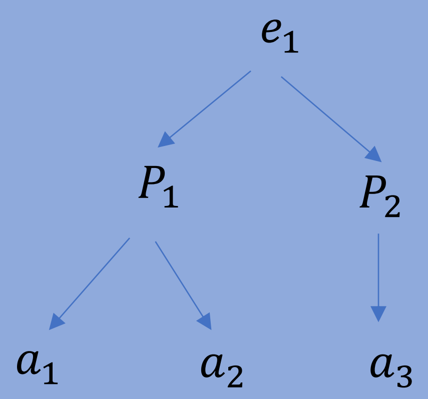
trace 2: e_1, P_2, a_3, length 3

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Observability : provide information of agent's progress of intentions

multiple plans



if the current step is at event e_1
what do we say about the progress?
by average: 1/((3+4)/2)= 1/3.5= 28.5%

trace 1: e_1, P_1, a_1, a_2, length 4

trace 2: e_1, P_2, a_3, length 3

each step
can be further
annotated
with different time units

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Directing Attention : direct attention to relevant environmental changes

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Directing Attention : direct attention to relevant environmental changes

Motivation rule: $\psi \rightsquigarrow \langle e, I \rangle \in \mathcal{M}$

Triggering condition ψ

an external event

motivation library

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Directing Attention : direct attention to relevant environmental changes

$$\psi \rightsquigarrow \langle e, I \rangle \in \mathcal{M}$$

1. allows the generation of multiple events based on one belief
   - $\psi \rightsquigarrow \langle e_1, I_1 \rangle, \cdots, \psi \rightsquigarrow \langle e_n, I_n \rangle$
2. benefits from the modularity principle by separating the following two
   - the dynamic of external event sets (i.e. desires);
   - the design of plan library.

Building Suitable Agents for Human-Autonomy Teaming

Observable and Attention-Directing BDI Agents

Directing Attention : direct attention to relevant environmental changes

$$\psi \rightsquigarrow \langle e, I \rangle \in \mathcal{M}$$

$$\frac{\psi \rightsquigarrow \langle e, I \rangle \in \mathcal{M} \qquad \mathcal{B} \vDash \psi \qquad \langle e, I \rangle \notin \Gamma}{\langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e \cup \langle e, I, active \rangle, \mathcal{B}, \Gamma \cup \{\langle P, I \rangle\} \rangle}$$

When an agent believes ψ, it should adopt the event $\langle e, I \rangle$ if it has not adopted it before.

Building Suitable Agents for Human-Autonomy Teaming

Executable Semantics for Observable and Attention-Directing BDI Agents

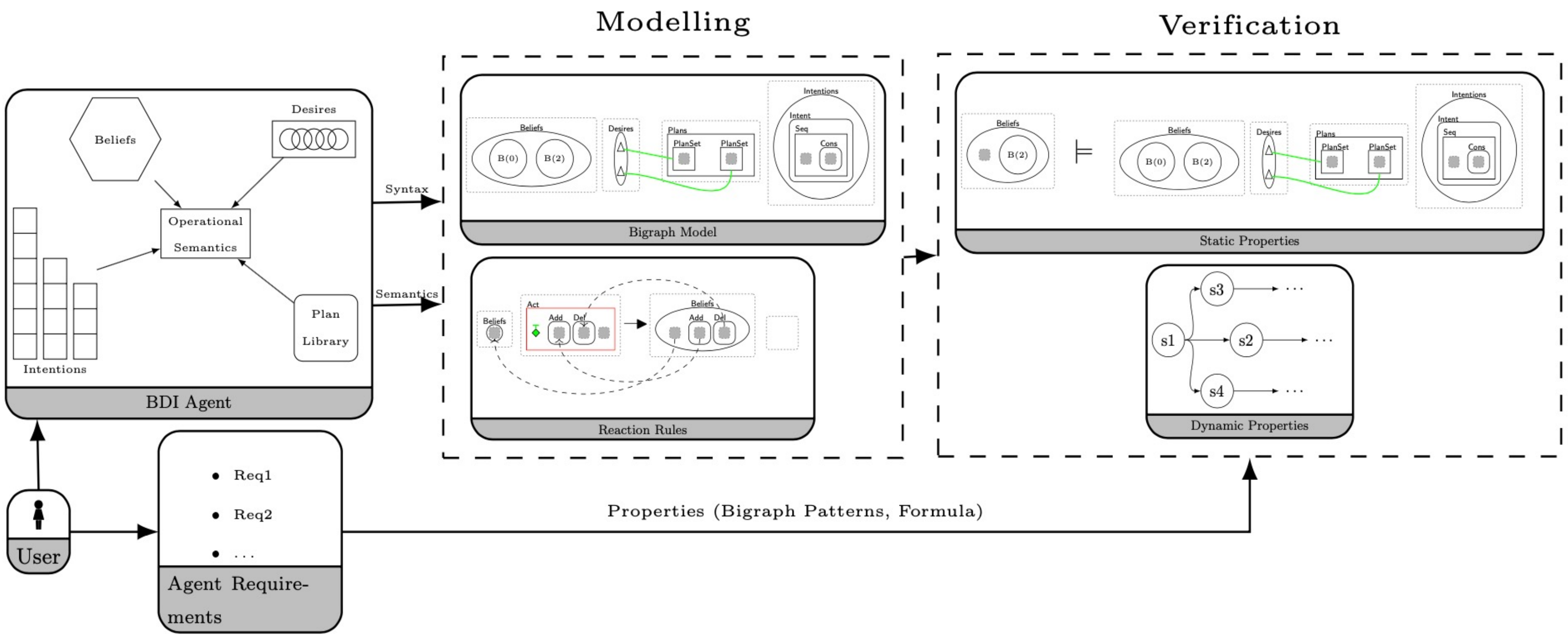


Figure 1: Modelling and verification framework for BDI agents.

Blair Archibald et al. (2021): Modelling and Verifying BDI Agents with Bigraphs. arXiv preprint arXiv:2105.02578 accepted in Science of Computer Programming

https://bitbucket.org/uog-bigraph/observable_attention-directing_bdi_model/src/master/

Building Suitable Agents for Human-Autonomy Teaming

Verification

Properties for observability:

1. if an intention is being progressed, it status should never be pending;
2. if an intention becomes a completed empty program, its related external event will eventually succeed;
3. if an intention becomes blocked, but is no an empty program, its related external event will eventually fail;

Building Suitable Agents for Human-Autonomy Teaming

Verification

Properties for observability:

1. if an intention is being progressed, it status should never be pending;
2. if an intention becomes a completed empty program, its related external event will eventually succeed;
3. if an intention becomes blocked, but is no an empty program, its related external event will eventually fail;

$A[\varphi_1 \Longrightarrow F(\varphi_2 \wedge \neg\varphi_3)]$ in Computation Tree Logic

bigraph patterns

$\varphi_1 \stackrel{\text{def}}{=} Intent_e.(Identifier.Identifier1 \mid Nil \mid id)$

$\varphi_3 \stackrel{\text{def}}{=} Event_e.(Identifier.Identifier1 \mid Failure \mid id)$

$\varphi_2 \stackrel{\text{def}}{=} Event_e.(Identifier.Identifier1 \mid Success \mid id)$

Building Suitable Agents for Human-Autonomy Teaming

Summaries of limitations

a preliminary work

Building Suitable Agents for Human-Autonomy Teaming

Summaries of limitations

a preliminary work

only focus one particular aspect of human-autonomy teaming

Question: what do you want the agent to tell you as it is working for or with you?

Building Suitable Agents for Human-Autonomy Teaming

Summaries of limitations

a preliminary work

only focus one particular aspect of human-autonomy teaming

Question: what do you want the agent to tell you as it is working for or with you?

Observability

Directing Attention

Building Suitable Agents for Human-Autonomy Teaming

Summaries of limitations

a preliminary work

only focus one particular aspect of human-autonomy teaming

Question: what do you want the agent to tell you as it is working for or with you?

Observability

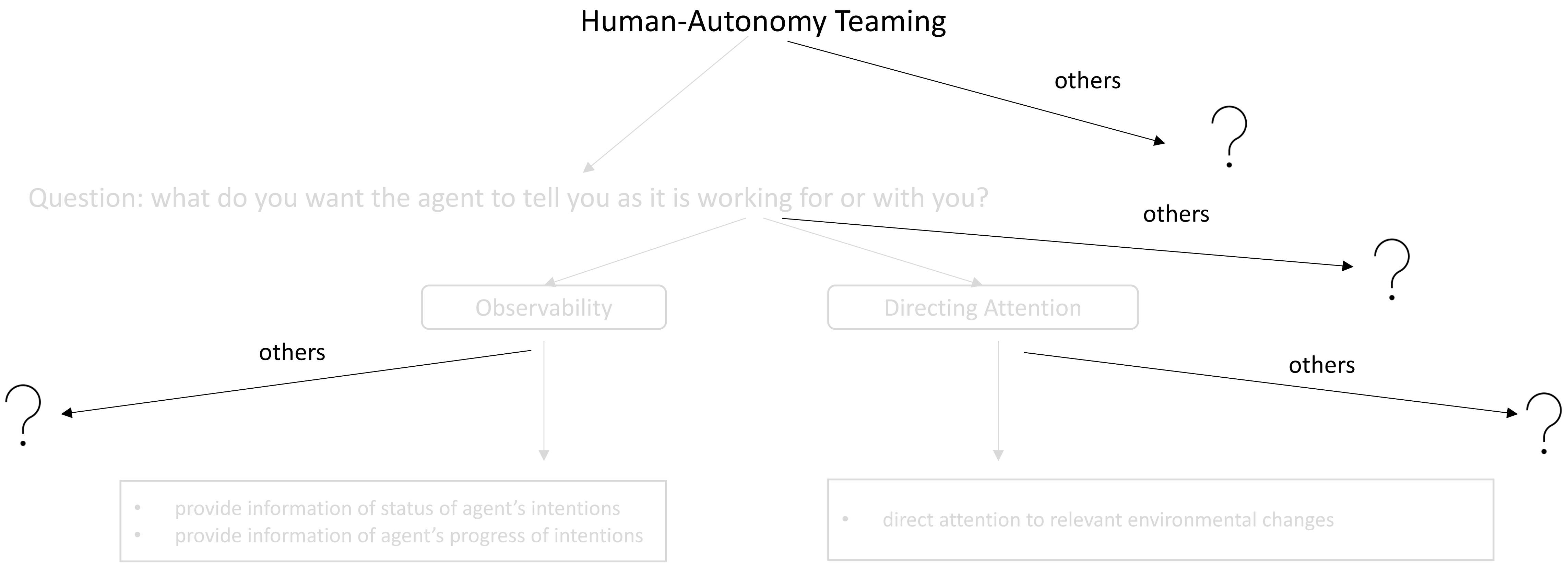
Directing Attention

- provide information of status of agent's intentions
- provide information of agent's progress of intentions

- direct attention to relevant environmental changes

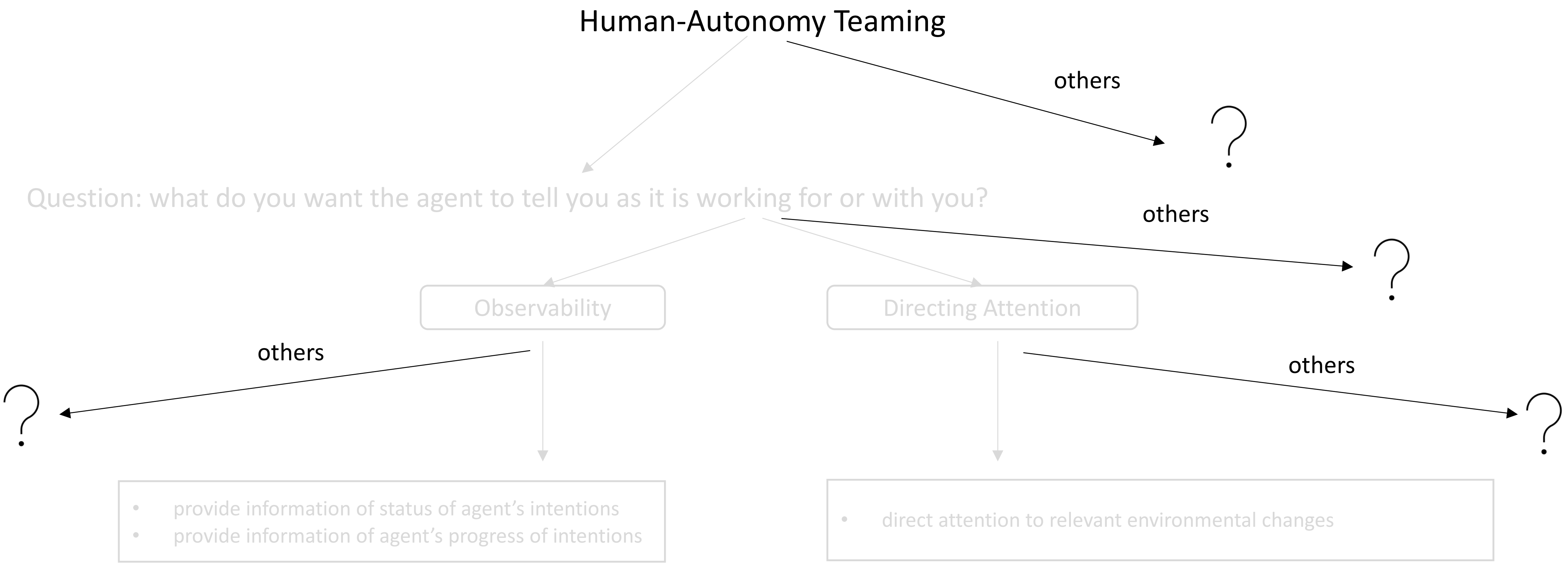
Building Suitable Agents for Human-Autonomy Teaming

Future work



Building Suitable Agents for Human-Autonomy Teaming

Future work



our bigraph-based executable semantics makes it easy to extend the model

Extensible Bigraph-based Executable BDI Model

Current work

Probabilistic BDI Agents: Actions, Plans, and Intentions

Blair Archibald, Muffy Calder, Michele Sevegnani, and Mengwei Xu

University of Glasgow, Glasgow, UK,
{blair.archibald, muffy.calder, michele.sevegnani, mengwei.xu}@glasgow.ac.uk

Abstract. The Belief-Desire-Intention (BDI) architecture is a popular framework for rational agents, yet most verification approaches are limited to analysing qualitative properties, for example whether an intention completes. BDI-based systems, however, operate in uncertain environments with dynamic behaviours: we may need quantitative analysis to establish properties such as the probability of eventually completing an intention. We define a probabilistic extension to the Conceptual Agent Notation (CAN) for BDI agents that supports probabilistic action outcomes, and probabilistic plan and intention selection. The semantics is executable via an encoding in Milner's bigraphs and the BigraphER tool. Quantitative analysis is conducted using PRISM. While the new semantics can be applied to any CAN program, we demonstrate the extension by comparing with standard plan and intention selection strategies (*e.g.* ordered or fixed schedules) and evaluating probabilistic action executions in a smart manufacturing scenario. The results show we can improve significantly the probability of intention completion, with appropriate probabilistic distribution. We also show the impact of probabilistic action outcomes can be marginal, even when the failure probabilities are large, due to the agent making smarter intention selection choices.

Keywords: BDI Agents; Quantitative Analysis; Bigraphs

to appear in SEFM'21 (25% acceptance rate)

Extensible Bigraph-based Executable BDI Model

Current work

Probabilistic BDI Agents: Actions, Plans, and Intentions

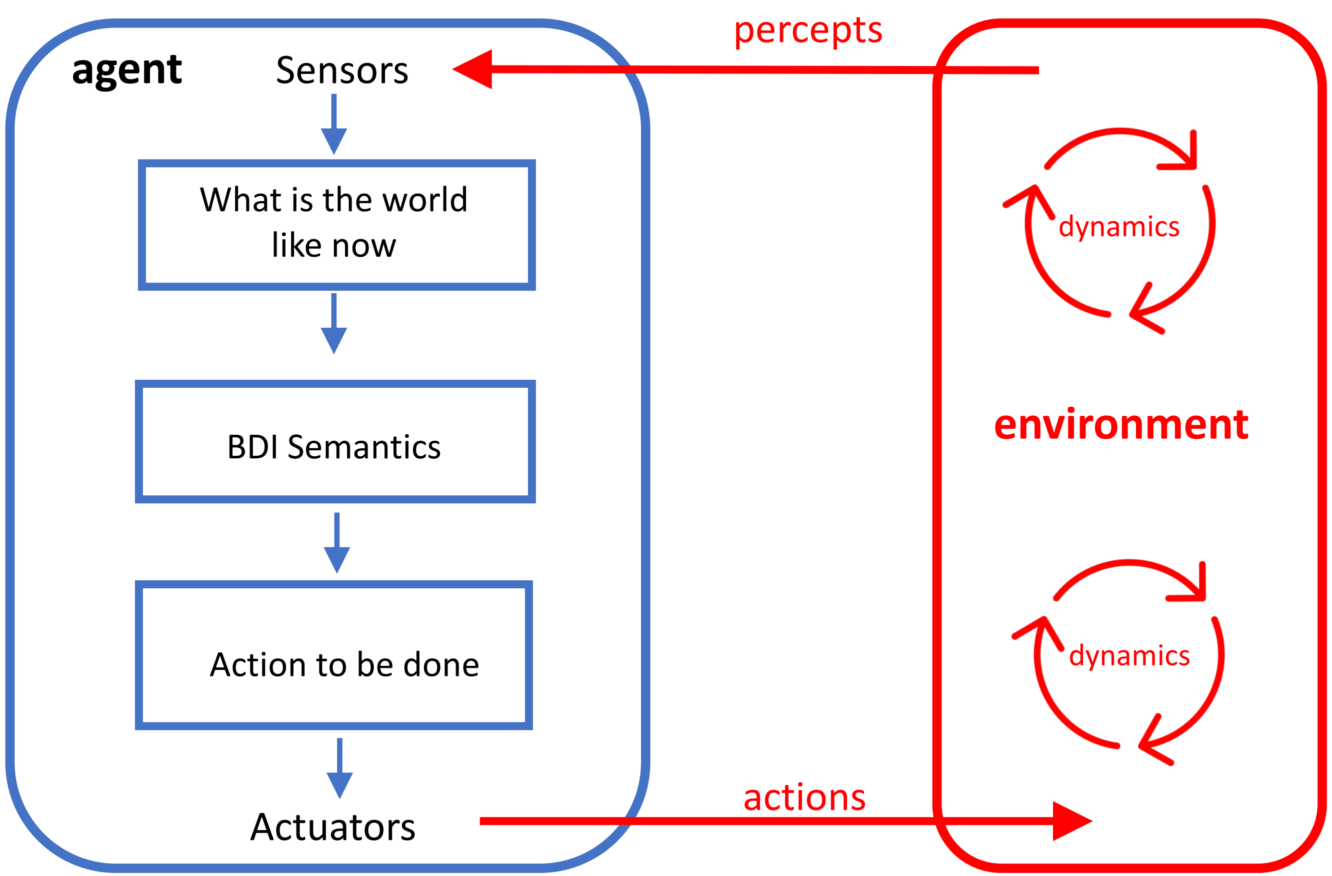
Blair Archibald, Muffy Calder, Michele Sevegnani, and Mengwei Xu

University of Glasgow, Glasgow, UK,
{blair.archibald, muffy.calder, michele.sevegnani,
mengwei.xu}@glasgow.ac.uk

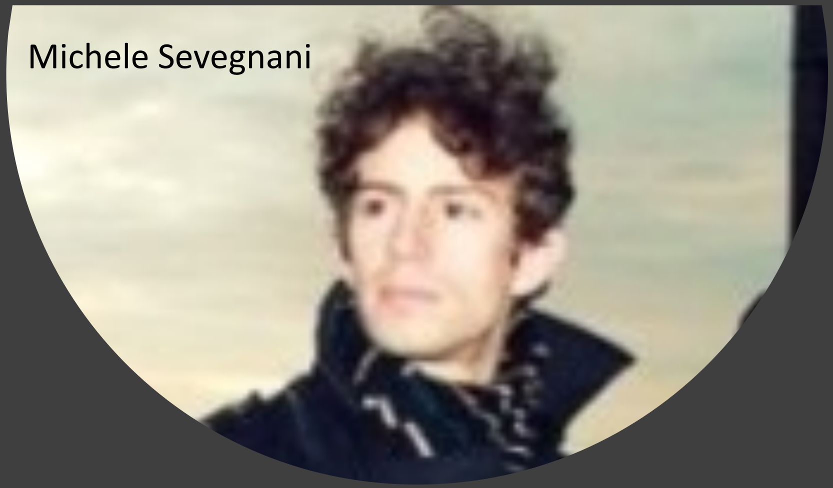
Abstract. The Belief-Desire-Intention (BDI) architecture is a popular framework for rational agents, yet most verification approaches are limited to analysing qualitative properties, for example whether an intention completes. BDI-based systems, however, operate in uncertain environments with dynamic behaviours: we may need quantitative analysis to establish properties such as the probability of eventually completing an intention. We define a probabilistic extension to the Conceptual Agent Notation (CAN) for BDI agents that supports probabilistic action outcomes, and probabilistic plan and intention selection. The semantics is executable via an encoding in Milner's bigraphs and the BigraphER tool. Quantitative analysis is conducted using PRISM. While the new semantics can be applied to any CAN program, we demonstrate the extension by comparing with standard plan and intention selection strategies (*e.g.* ordered or fixed schedules) and evaluating probabilistic action executions in a smart manufacturing scenario. The results show we can improve significantly the probability of intention completion, with appropriate probabilistic distribution. We also show the impact of probabilistic action outcomes can be marginal, even when the failure probabilities are large, due to the agent making smarter intention selection choices.

Keywords: BDI Agents; Quantitative Analysis; Bigraphs

to appear in SEFM'21 (25% acceptance rate)



Michele Sevegnani



Blair Archibald

Mengwei Xu

Muffy Calder

Many thanks for your attentions

University of Glasgow