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)
Human Autonomy Teaming

Definition:

A term describing

humans,

autonomous agents,

working together to achieve some objectives

Human Autonomy Teaming

Building suitable agents for human-autonomy teaming:

A term describing humans, autonomous agents, working together to achieve some objectives.

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?

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:

**Logics**
- [Cohen & Levesque, 1990]
- [Rao & Georgeff, 1991]
- [Shoham, 2009]
- ...

**Programming Languages**
- AgentSpeak [Rao, 1996]
- CAN [Winikoff et al., 2002]
- 3APL [Dastani et al., 2005]
- ...

**Software Platforms**
- Jason [Bordini et al., 2007]
- Jack [Winikoff, 2005]
- Jadex [Pokahr et al., 2013]
- ...

...
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:

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**
  - provide information of status of agent’s intentions
  - provide information of agent’s progress of intentions

- **Directing Attention**
  - direct attention to relevant environmental changes
Observability: provide information of status (e.g., success/failure) of agent’s intentions

External Event: $e \in E^e$

  e.g. new goal
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: $\langle e, I, \text{status} \rangle \in E^e$

- the unique identifier
- $\text{status} \in \{\text{pending, active, success, fail}\}$
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

\[
\begin{align*}
& e \in E^e \\
& \langle E^e, \mathcal{B}, \Gamma \rangle \Rightarrow \langle E^e \setminus \{e\}, \mathcal{B}, \Gamma \cup \{e\} \rangle
\end{align*}
\]

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:

\[
\begin{align*}
\forall e \in E^e, \quad & \langle E^e, B, \Gamma \rangle \Rightarrow \langle E^e \setminus \{e\}, B, \Gamma \cup \{e\} \rangle \\
& A_{\text{event}}
\end{align*}
\]

New CAN semantics:

\[
\begin{align*}
\forall (e, I, \text{pending}) \in E^e, \quad & \langle E^e, B, \Gamma \rangle \Rightarrow \langle E^e \setminus \{(e, I, \text{pending})\} \cup \{e, I, \text{active}\}, B, \Gamma \cup \{(e, I)\} \rangle \\
& A_{\text{new event}}
\end{align*}
\]

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 Update**

\[
\begin{align*}
P & \in \Gamma \\
\langle B, P \rangle & \not\rightarrow \\
\langle E^e, B, \Gamma \rangle & \Rightarrow \langle E^e, B', \Gamma \setminus \{P\} \rangle
\end{align*}
\]

**New Update**

**Success**

\[
\begin{align*}
(P, I) & \in \Gamma \\
\langle e, I, active \rangle & \in E^e \\
\langle B, \langle P, I \rangle \rangle & \not\rightarrow \\
(P = \text{nil}) & \\
\langle E^e, B, \Gamma \rangle & \Rightarrow \langle E^e \setminus \{(e, I, active)\} \cup \langle e, I, success \rangle, B, \Gamma \setminus \{\langle P, I \rangle\} \rangle
\end{align*}
\]

**Failure**

\[
\begin{align*}
(P, I) & \in \Gamma \\
\langle e, I, active \rangle & \in E^e \\
\langle B, \langle P, I \rangle \rangle & \not\rightarrow \\
(P \neq \text{nil}) & \\
\langle E^e, B, \Gamma \rangle & \Rightarrow \langle E^e \setminus \{(e, I, active)\} \cup \langle e, I, failure \rangle, B, \Gamma \setminus \{\langle P, I \rangle\} \rangle
\end{align*}
\]
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

\[ \text{plan } P_1 = e_1: \varphi_1 \leftarrow a_1; a_2 \]
Building Suitable Agents for Human-Autonomy Teaming
Observable and Attention-Directing BDI Agents

Observability: provide information of agent’s progress of intentions

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

intuition

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**
  - 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
  - $e_1$
  - $P_1$
  - $a_1$
  - $a_2$
  - $P_2$
  - $a_3$

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

![Diagram](Image)

- Multiple plans
  - Event $e_1$
  - Plans $P_1$, $P_2$
  - Actions $a_1$, $a_2$, $a_3$

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: \( \psi \implies (e, I) \in \mathcal{M} \)

- Triggering condition \( \psi \)
- an external event
- motivation library
Directing Attention: direct attention to relevant environmental changes

\[ \psi \rightsquigarrow (e, I) \in \mathcal{M} \]

1. allows the generation of multiple events based on one belief
   - \[ \psi \rightsquigarrow (e_1, I_1), \ldots, \psi \rightsquigarrow (e_n, I_n) \]
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.
Directing Attention: direct attention to relevant environmental changes

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

\[
\begin{align*}
\psi \rightsquigarrow \langle e, I \rangle \in \mathcal{M} &\quad \mathcal{B} \models \psi \quad \langle e, I \rangle \notin \Gamma \\
\left\langle E^e, \mathcal{B}, \Gamma \right\rangle &\Rightarrow \left\langle E^e \cup \langle e, I, \text{active} \rangle, \mathcal{B}, \Gamma \cup \{\langle P, I \rangle\} \right\rangle
\end{align*}
\]

When an agent believes \( \psi \), 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.


https://bitbucket.org/uog-bigraph/observable_attention-directing_bdi_model/src/master/
Properties for observability:
1. if an intention is being progressed, its 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;
Properties for observability:
1. if an intention is being progressed, its 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;

\[ A[\varphi_1 \Rightarrow F(\varphi_2 \land \neg \varphi_3)] \text{ in Computation Tree Logic} \]

\[ \varphi_1 \equiv \text{Intent}_e. (\text{Identifier}.\text{Identifier1} \mid \text{Nil} \mid \text{id}) \]

\[ \varphi_2 \equiv \text{Event}_e. (\text{Identifier}.\text{Identifier1} \mid \text{Success} \mid \text{id}) \]

\[ \varphi_3 \equiv \text{Event}_e. (\text{Identifier}.\text{Identifier1} \mid \text{Failure} \mid \text{id}) \]
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

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

- Observability
- Directing Attention

a preliminary work

only focus one particular aspect of human-autonomy teaming
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

• provide information of status of agent’s intentions
• provide information of agent’s progress of intentions

Directing Attention

• direct attention to relevant environmental changes
Building Suitable Agents for Human-Autonomy Teaming

Future work

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

**Human-Autonomy Teaming**

- Observability
  - others
  - provide information of status of agent’s intentions
  - provide information of agent’s progress of intentions

- Directing Attention
  - others
  - direct attention to relevant environmental changes
Building Suitable Agents for Human-Autonomy Teaming

Future work

Human-Autonomy Teaming

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

Observability

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

Directing Attention

- direct attention to relevant environmental changes

other

our bigraph-based executable semantics makes it easy to extend the model
Probabilistic BDI Agents: Actions, Plans, and Intentions

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

University of Glasgow, Glasgow, UK,
{blair.archibald, muffy.calder, michele.sevignani, 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 analyzing 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 schedule) 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)
Many thanks for your attentions