Verifying Autonomous Agents in Dynamic Environment

Mengwei Xu

University of Glasgow

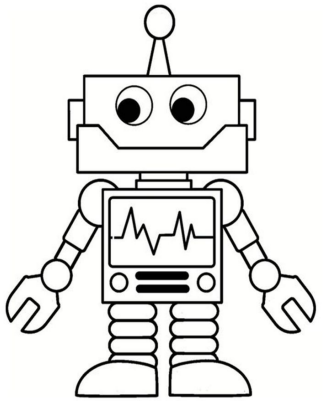
Autonomous Agents

Definition:

An entity

which **perceives** its environment,
which **deliberates** accordingly,
which **takes actions** autonomously,

in order to achieve some objectives



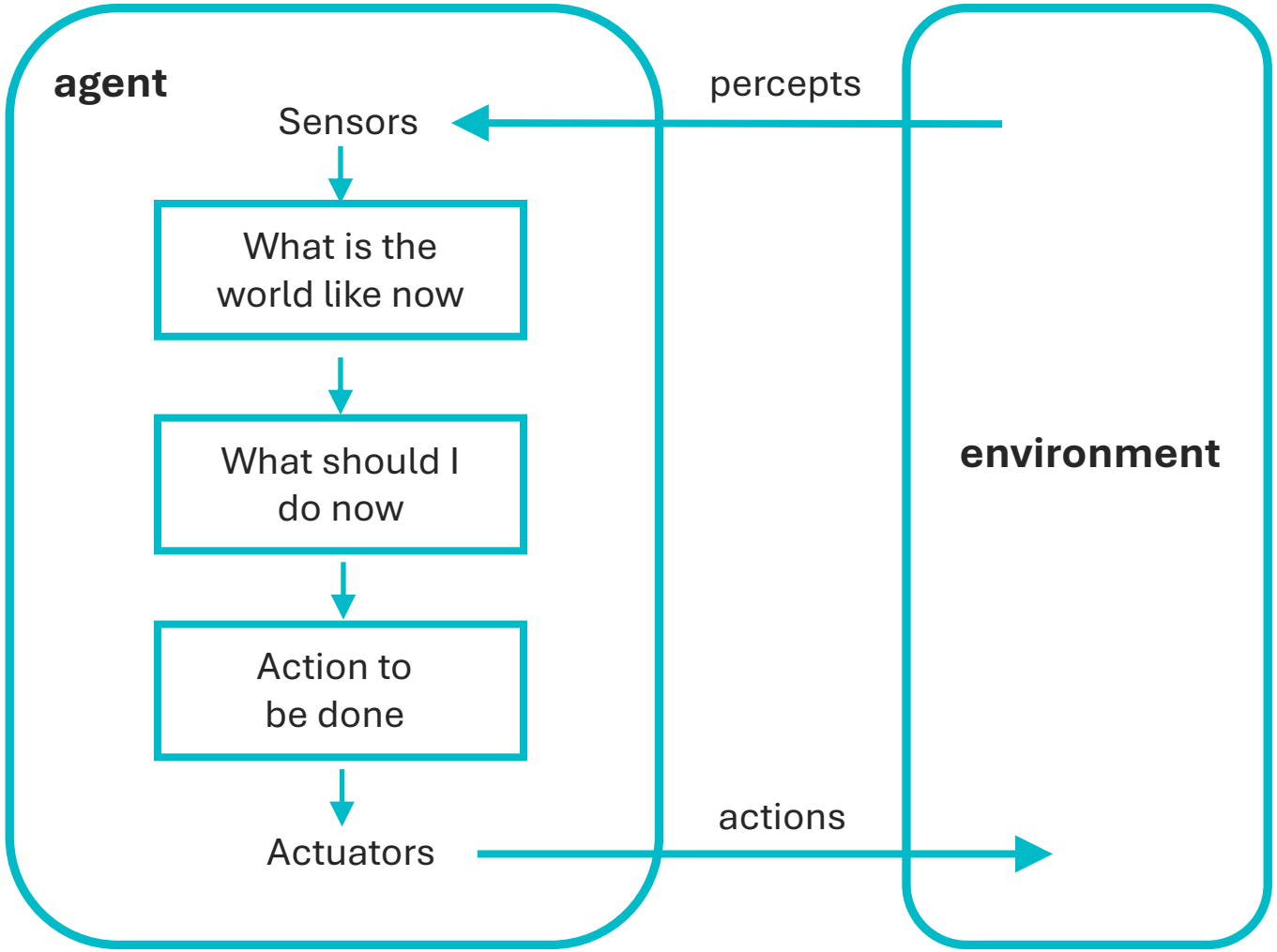
Autonomous Agents

Definition

An entity

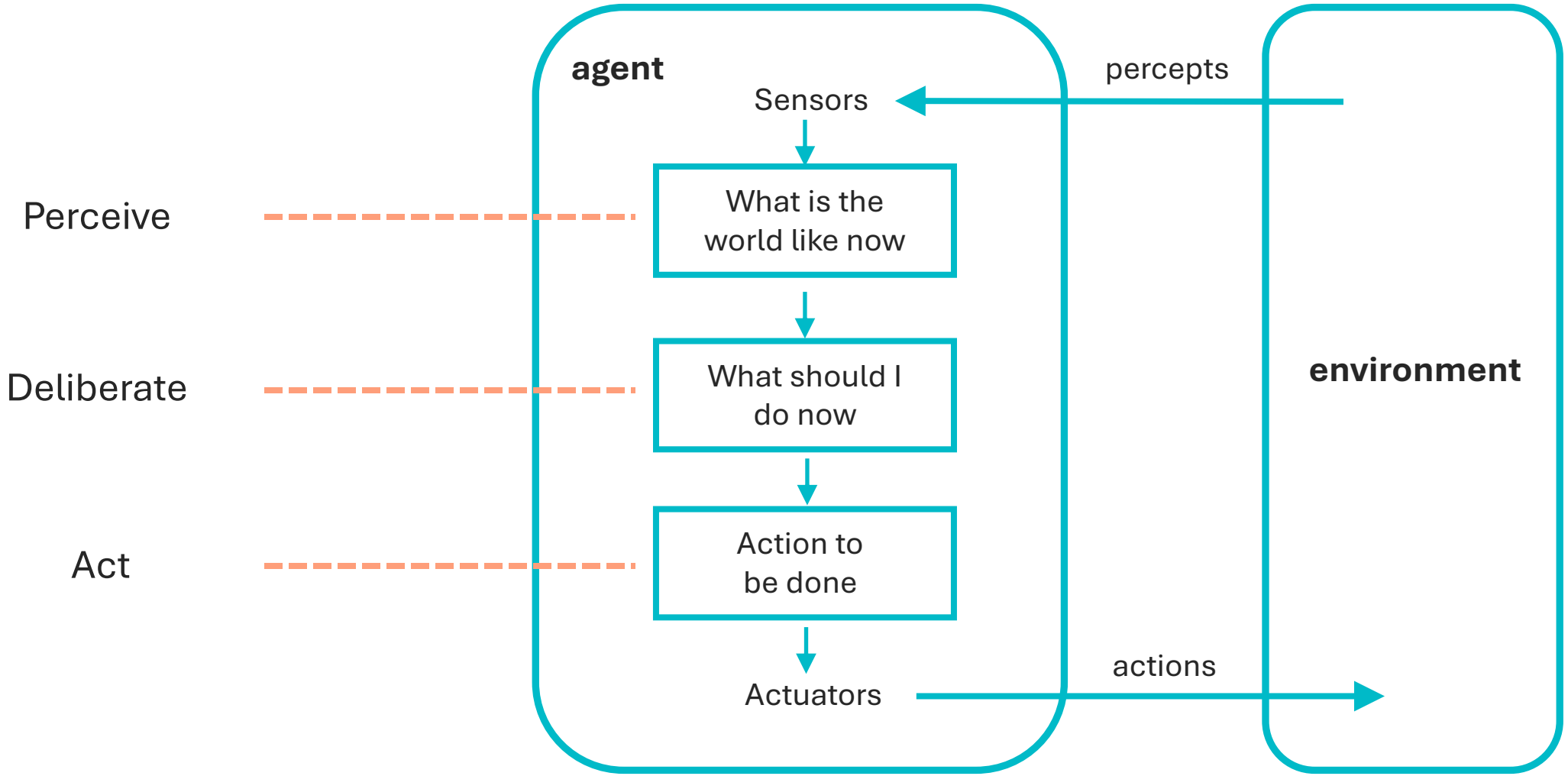
which **perceives** its environment,
which **deliberates** accordingly,
which **takes actions** autonomously,

in order to achieve some objectives



Autonomous Agents

Reasoning Cycle



Autonomous Agents

Reasoning Cycle

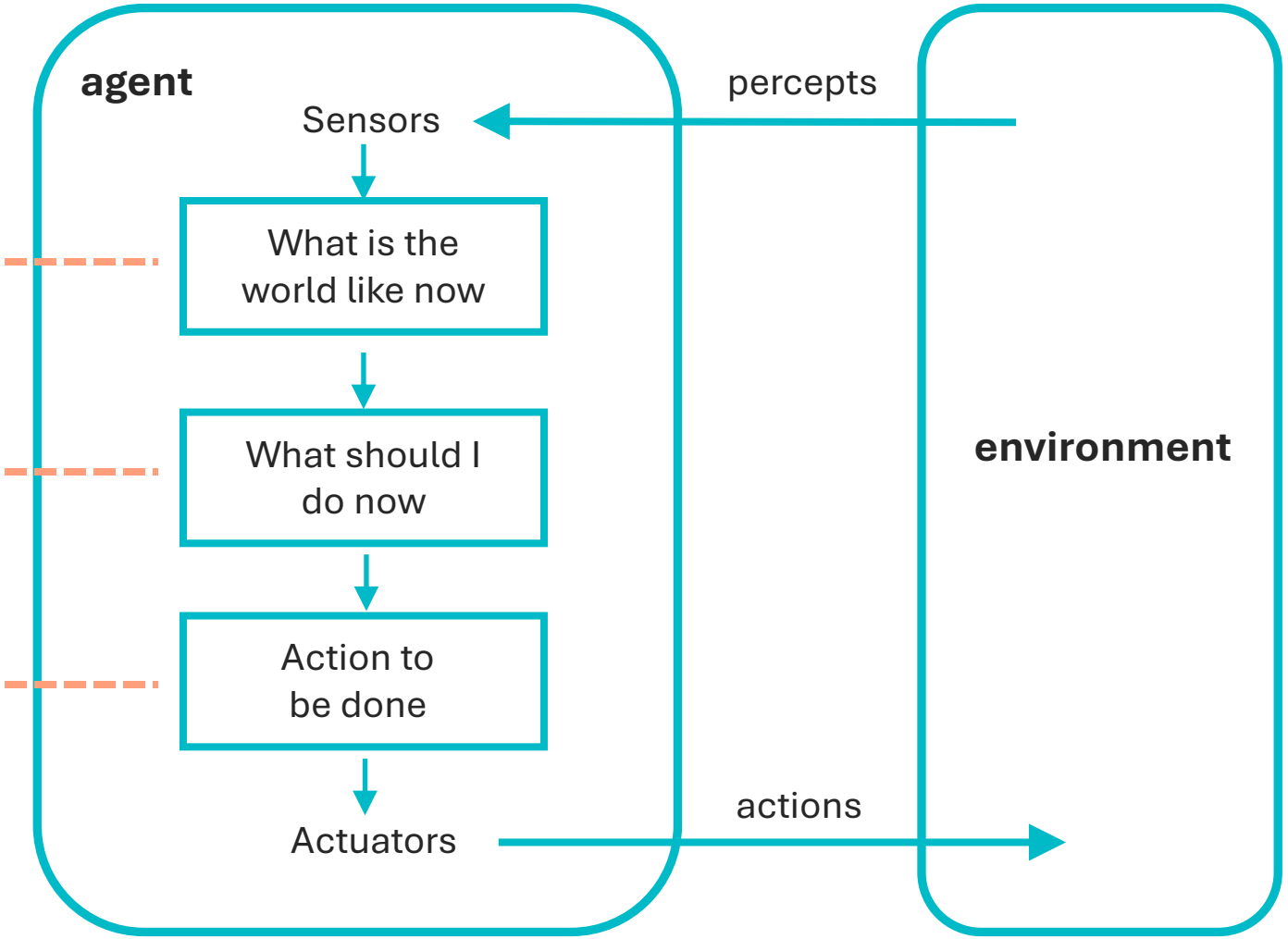
while true:

Perceive

Deliberate

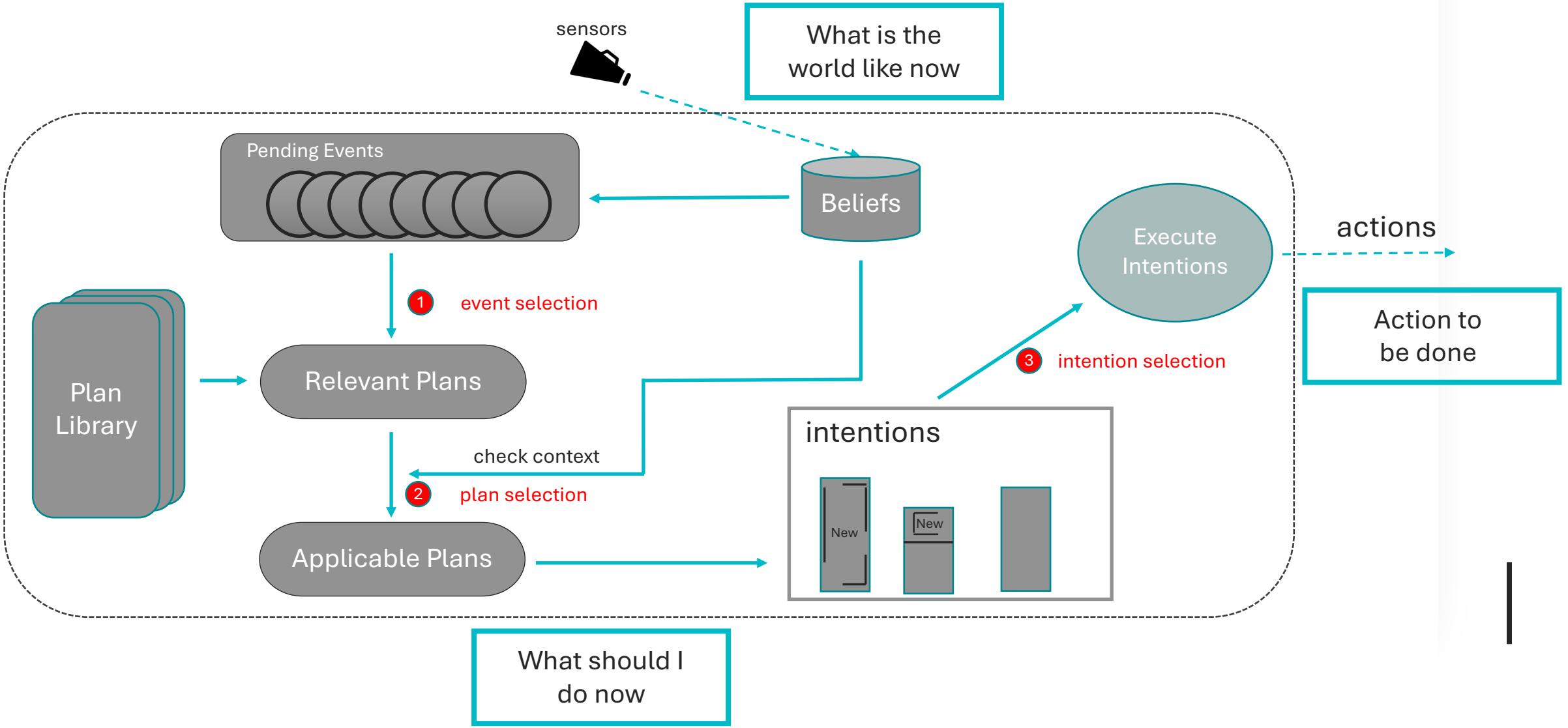
Act

end



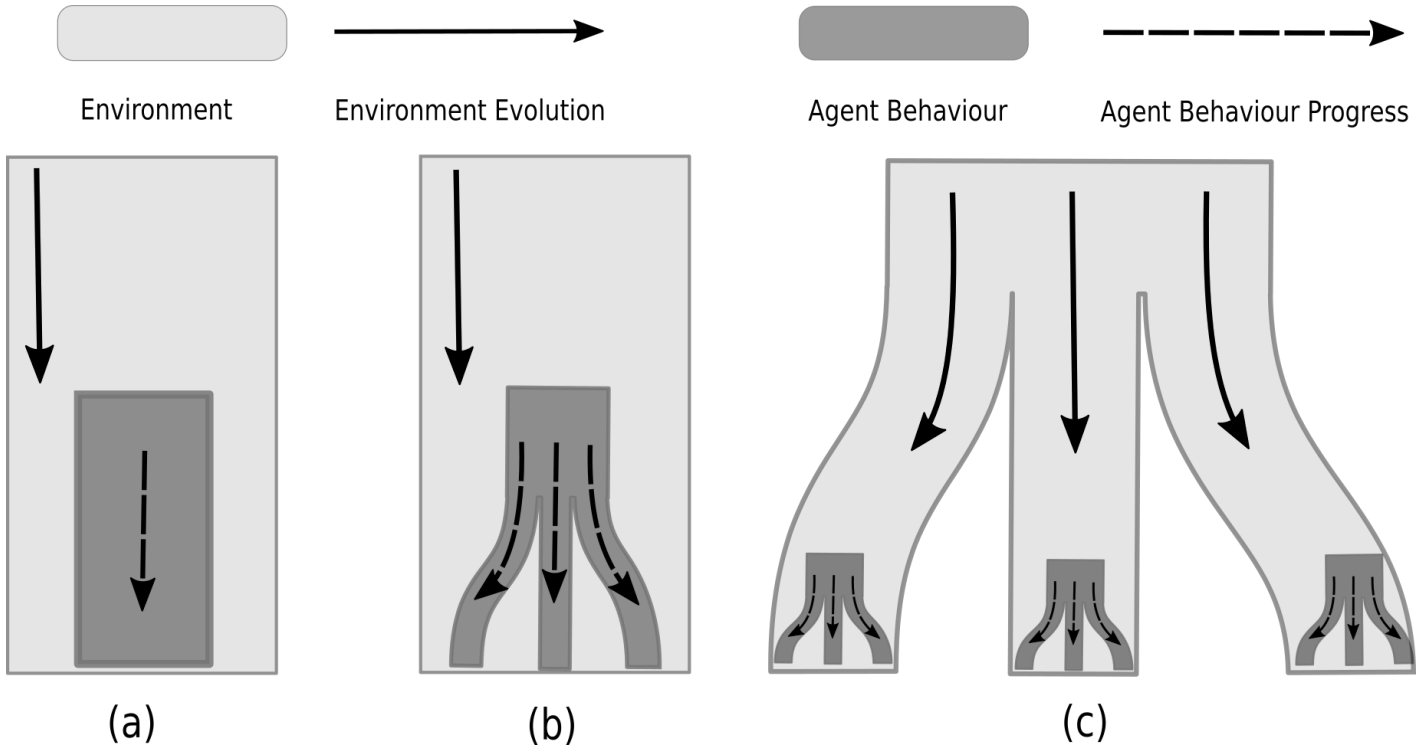
Autonomous Agents

Beliefs-Desires-Intentions (BDI) Framework



Autonomous Agents

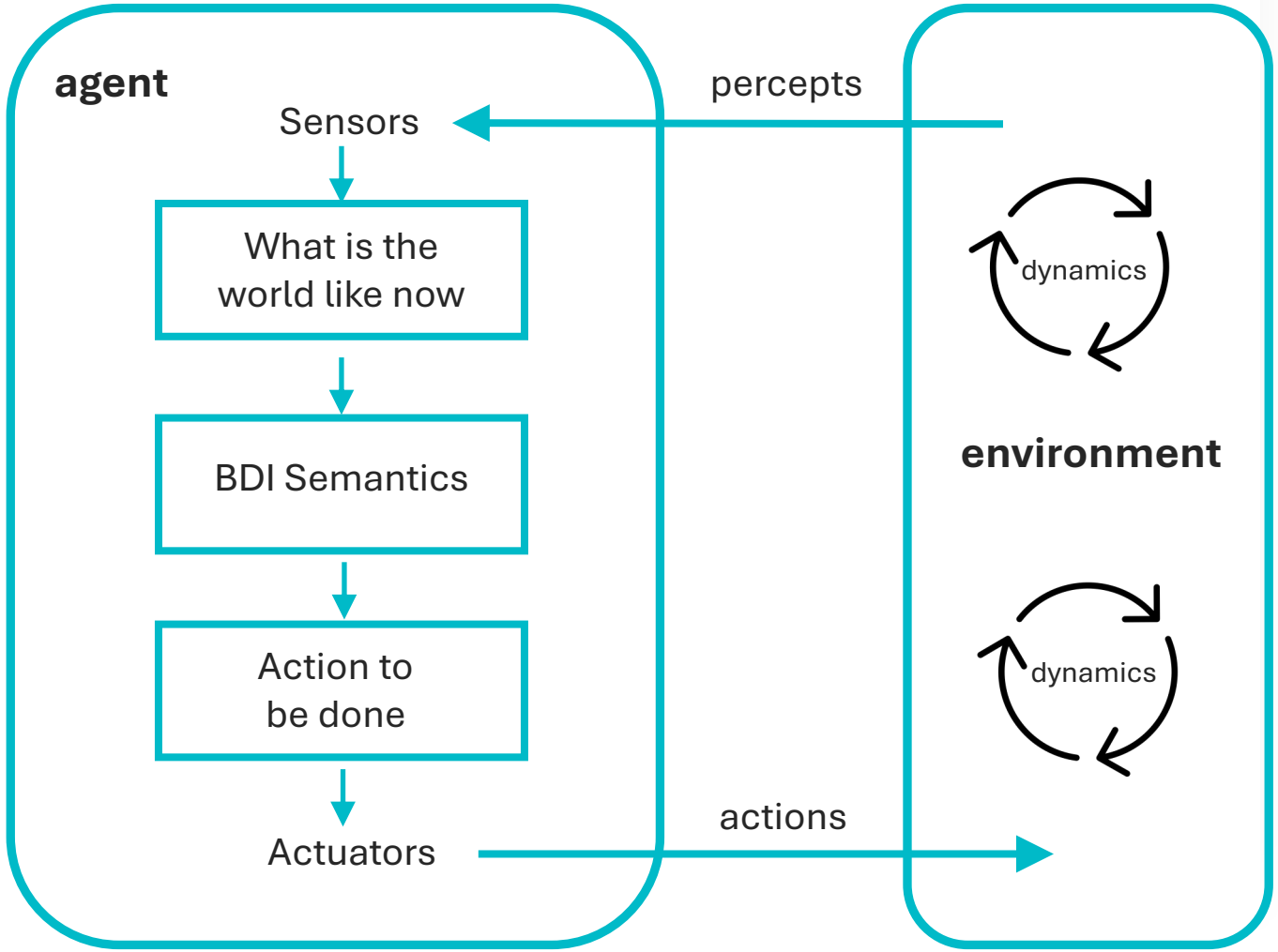
Related Work



(a) simulation: one run of agent behaviour in one environment;
(b) existing verification approaches: all possible agent behaviours in one environment
(c) our proposed approach: verify all possible agent behaviours in all possible environments

Autonomous Agents

Verification Framework



Autonomous Agents

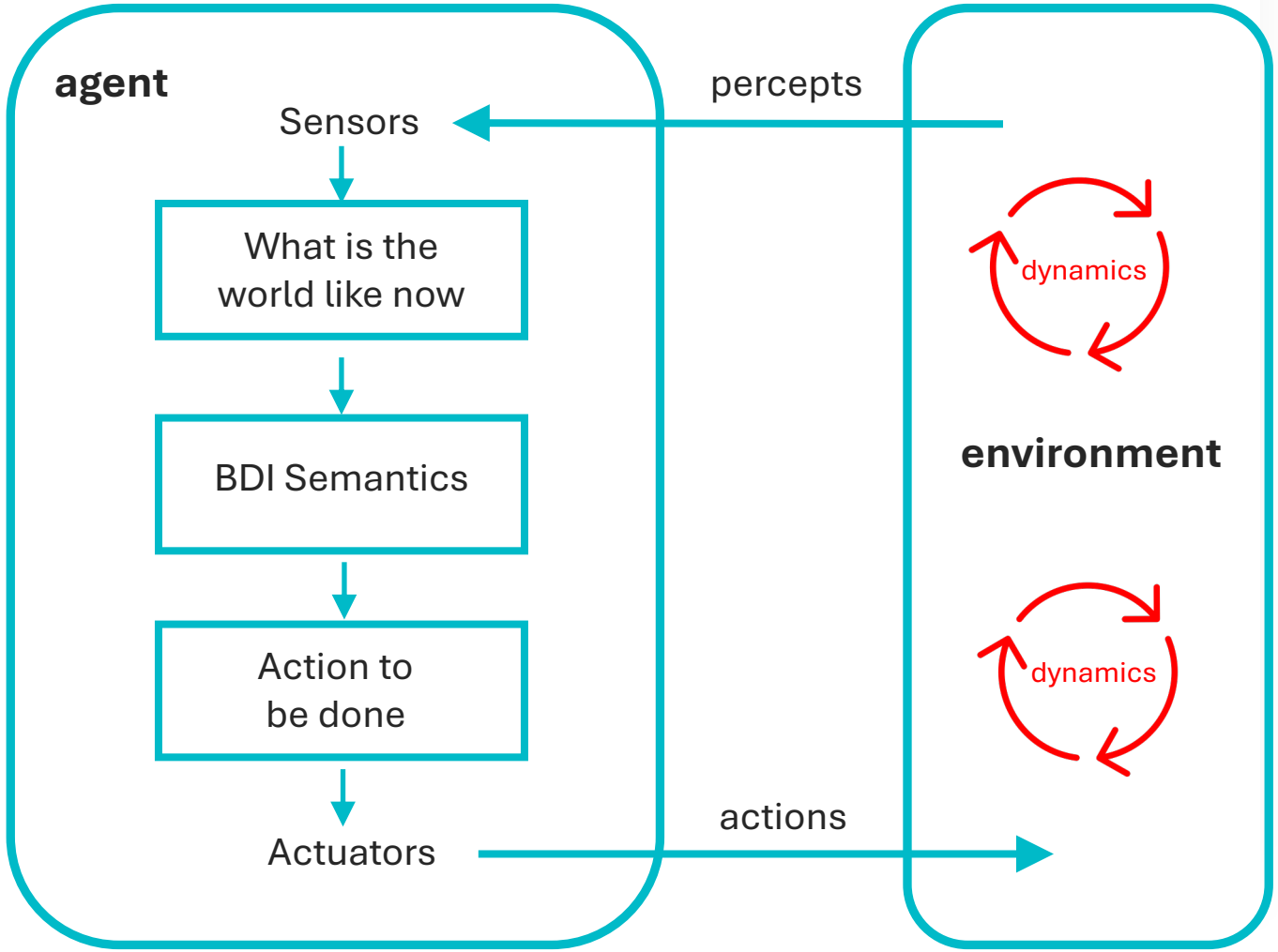
Verification Framework

normal environment changes
such as from p to ¬ p

while true:

environment update

end



Autonomous Agents

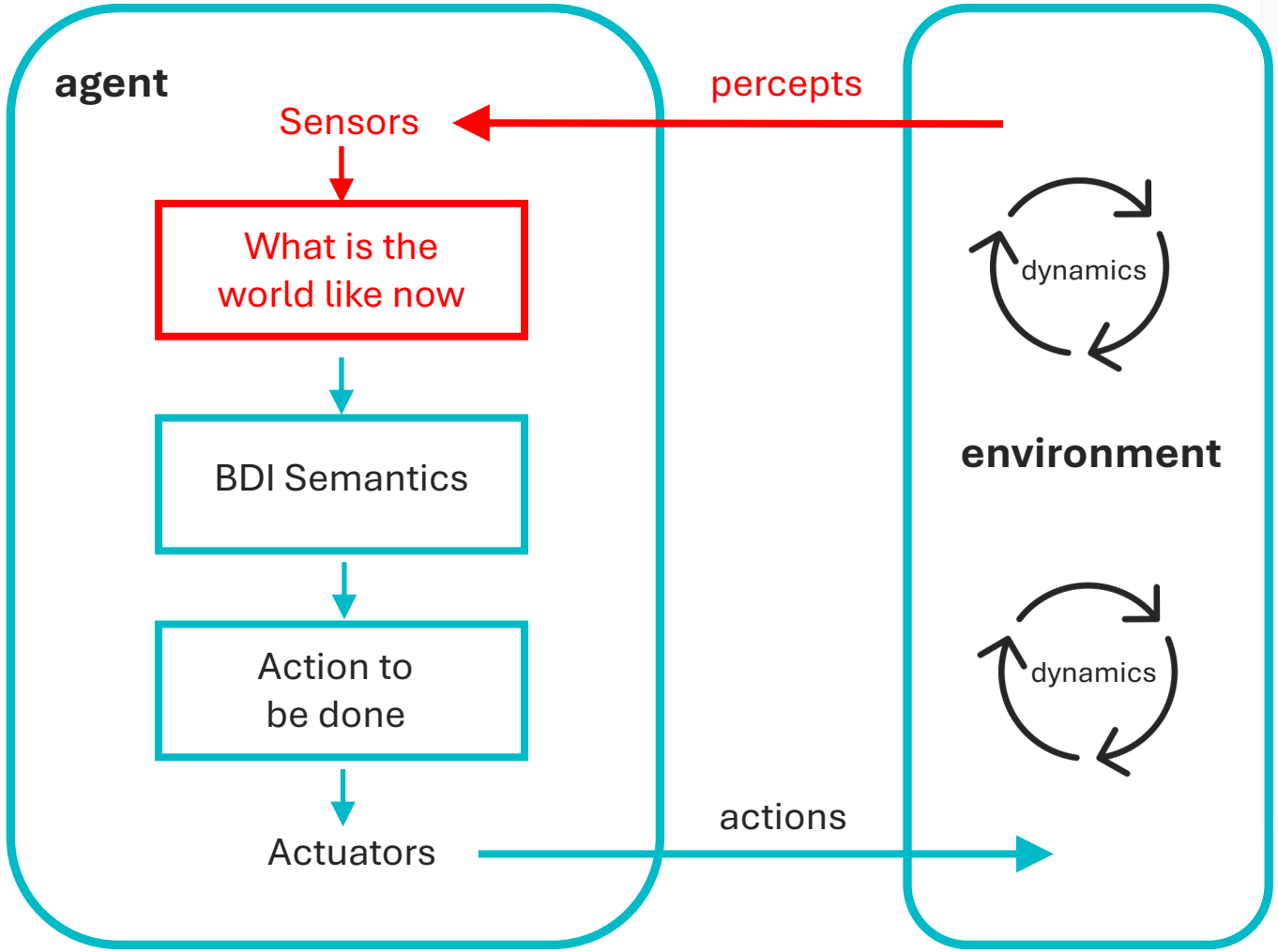
Verification Framework

while true:

environment update

perceive

end



Autonomous Agents

Verification Framework

while true:

environment update

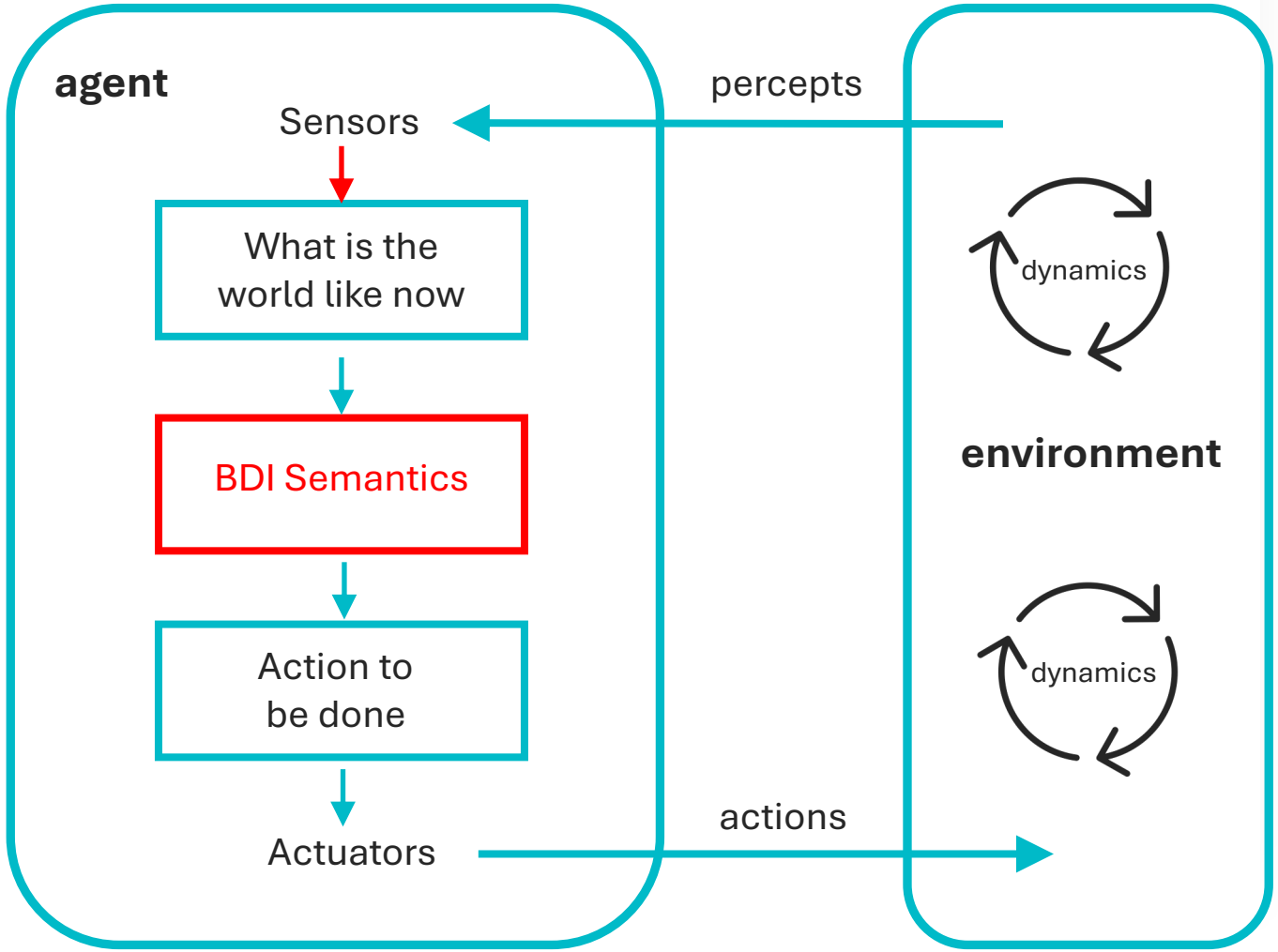
perceive

while true:

one agent semantic step

end

end



Autonomous Agents

Verification Framework

while true:

environment update

perceive

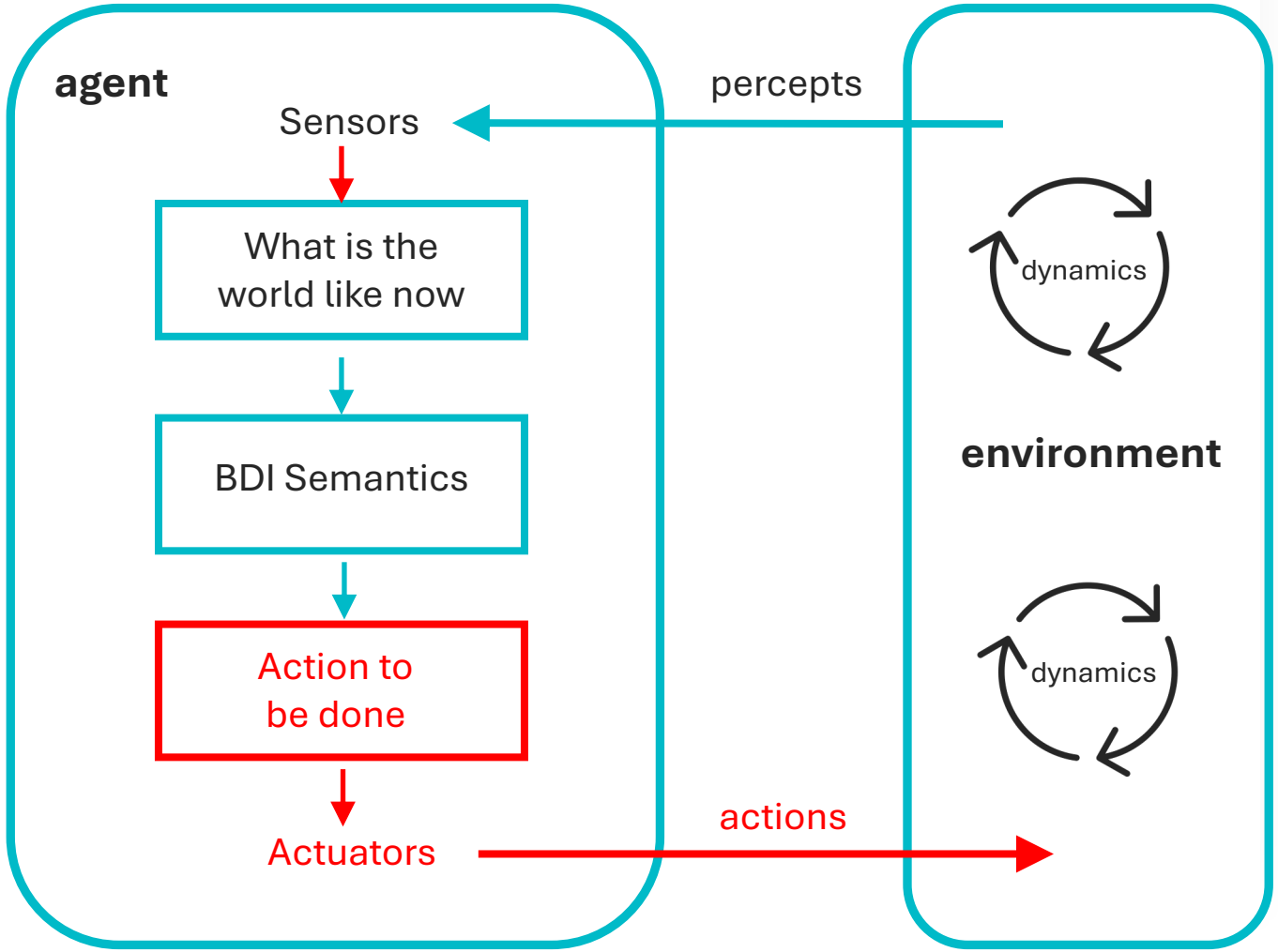
while true:

one agent semantic step

end

act

end



Autonomous Agents

Executable Framework

while true:

environment update

perceive

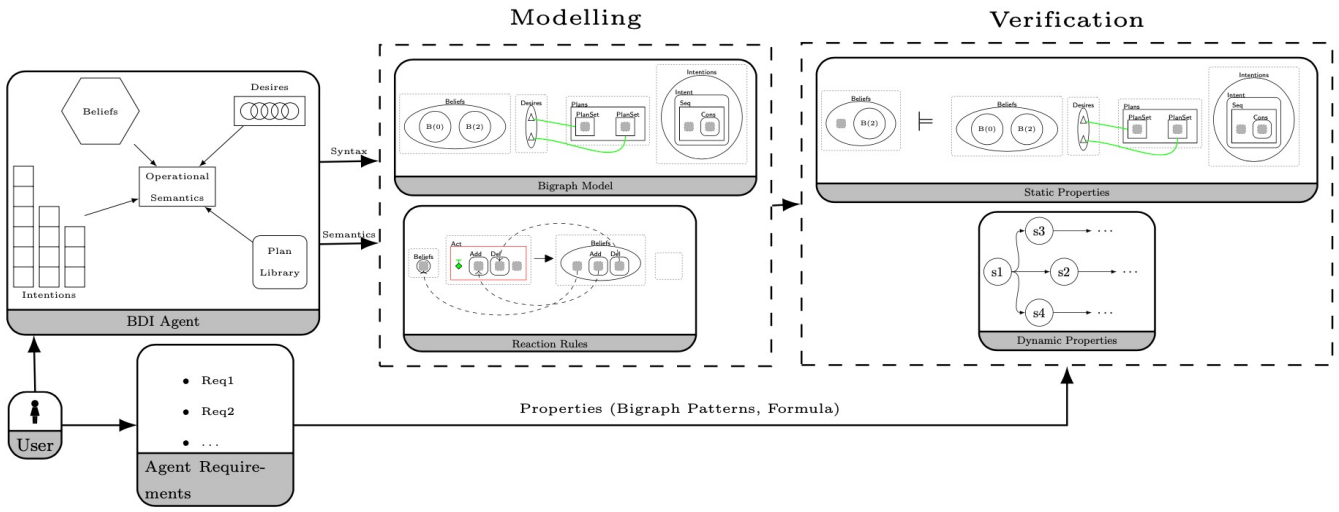
while true:

one agent semantic step

end

act

end



Contents lists available at ScienceDirect

Science of Computer Programming

www.elsevier.com/locate/scico

ELSEVIER

Modelling and verifying BDI agents with bigraphs

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

School of Computing Science, University of Glasgow, UK

Autonomous Agents

Examples

1 **Plan library**

2 e_patrol_init : true ← goal(detection, e_patrol_task, false); return

3 e_patrol_task : true ← goal(harsh_weather, e_patrol, false); e_pause

4 e_patrol : true ← patrol

5 e_pause : harsh_weather ∧ ¬parked ← activate_parking; wait

6 e_pause : harsh_weather ∧ parked ← wait

7 **initial environment state**

8 $\Theta_0 = \{\neg a, \neg b, \neg c, \neg d, e\_patrol\_init\}$

9 **environment transition function**

$$\delta(\Theta) = \begin{cases} \{\Theta, (\Theta \setminus \{\neg a\}) \cup \{a\}, (\Theta \setminus \{\neg b\}) \cup \{b\}, (\Theta \setminus \{\neg a, \neg b\}) \cup \{a, b\}\} & \text{if } \neg a \wedge \neg b \in \Theta \quad (1) \\ \{\Theta, (\Theta \setminus \{\neg a\}) \cup \{a\}\}, & \text{if } \neg a \wedge b \in \Theta \quad (2) \\ \{\Theta, (\Theta \setminus \{\neg b\}) \cup \{b\}\}, & \text{if } a \wedge \neg b \in \Theta \quad (3) \\ \{\Theta\}, & \text{if } a \wedge b \in \Theta \quad (4) \\ \{(\Theta \setminus \{b, c\}) \cup \{\neg b, \neg c\}\}, & \text{if } b \wedge c \in \Theta \quad (5) \end{cases}$$

where a = detection, b = harsh_weather, c = waited (the effect of action wait) and d = returned (the effect of action return).

agent design

possible environment changes

Autonomous Intelligent Agents

Examples

	Design in Fig. 5	Design in Fig. 6
Saftey Property	False	True
Completion Property	False	False
Response Property	True	True
Commitment Property	True	True
States	167	282
Transitions	242	373
Build time (s)	54.05	128.89
Rule applications	1306	2152

Table I: Properties checked: where safety property is $\neg\mathbf{E}[\mathbf{F}(\varphi_1 \wedge \neg\varphi_2 \wedge (\mathbf{XX}\varphi_2))]$, completion property $\mathbf{A}[\mathbf{F}\varphi_3]$, response property $\mathbf{A}[\varphi_4 \implies \mathbf{F}\varphi_5]$, and commitment property $\mathbf{A}[\varphi_5 \implies \mathbf{F}\varphi_6]$.

φ_1 = harsh_weather φ_2 = returned

Autonomous Agents

Future Work

while true:

environment update

perceive

while true:

one agent semantic step

end

act

end

1. normal environment changes such as from p to not p
2. the request of new events
3. the command of event status changes

yet to be implemented

Questions