An Agent-Centric Perspective on Norm Enforcement and Sanctions

Elena Yan* Luis G. Nardin* Jomi F. Hübner** Olivier Boissier*

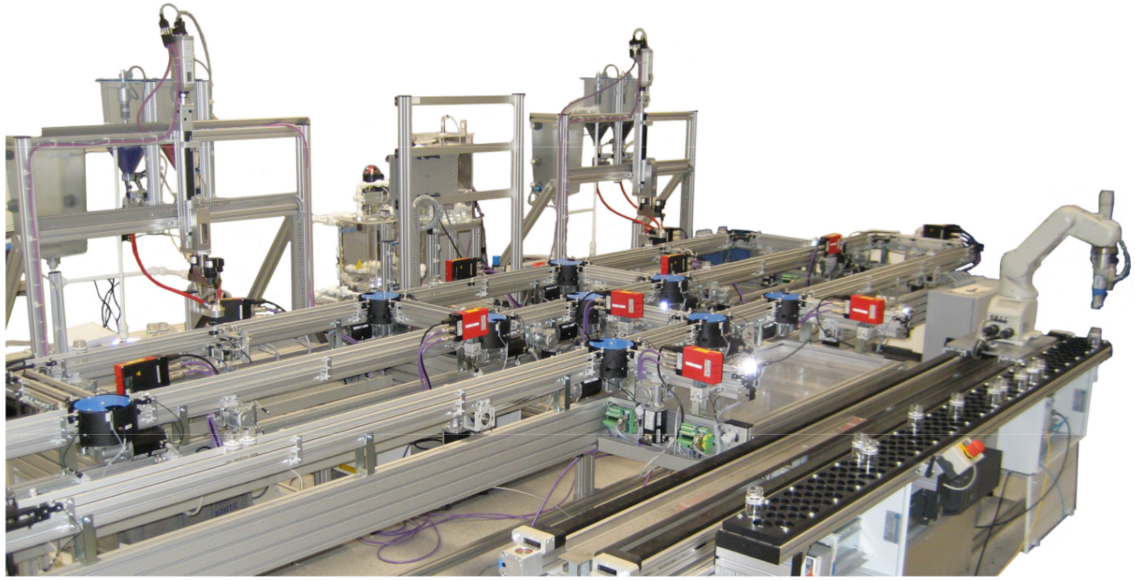
*MINES Saint-Etienne, UMR 6158 LIMOS, France
{elena.yan, luisgustavo.nardin, olivier.boissier}@emse.fr

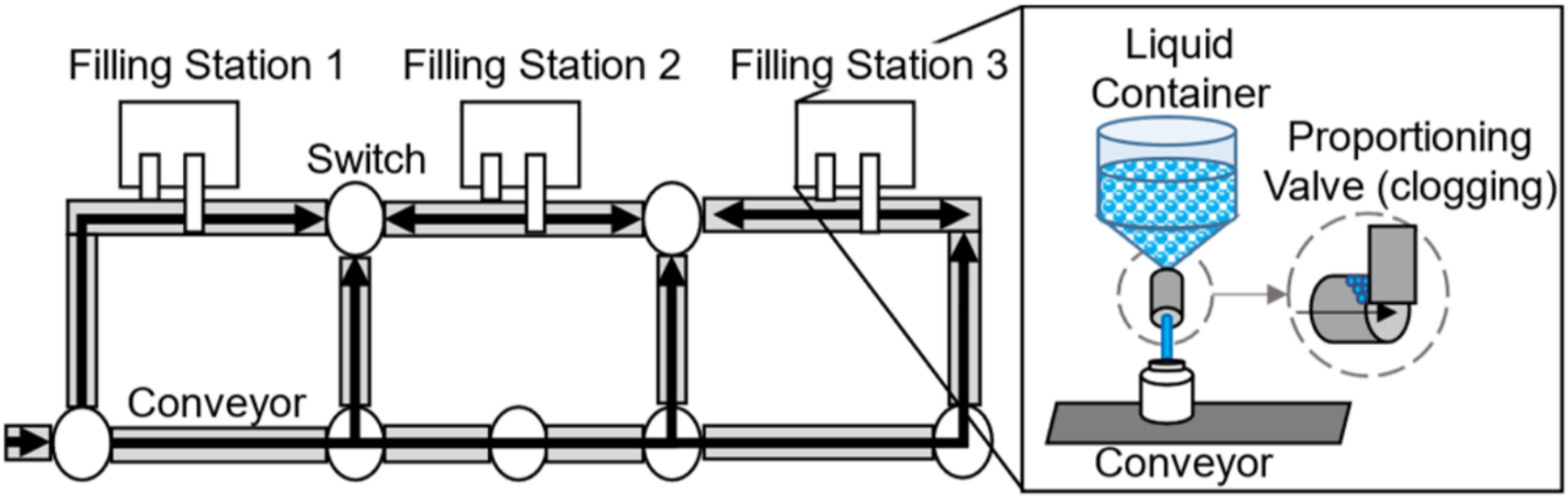
**Federal University of Santa Catarina, Florianópolis, Brazil
jomi.hubner@ufsc.br

WOA 2024, July 9, 2024

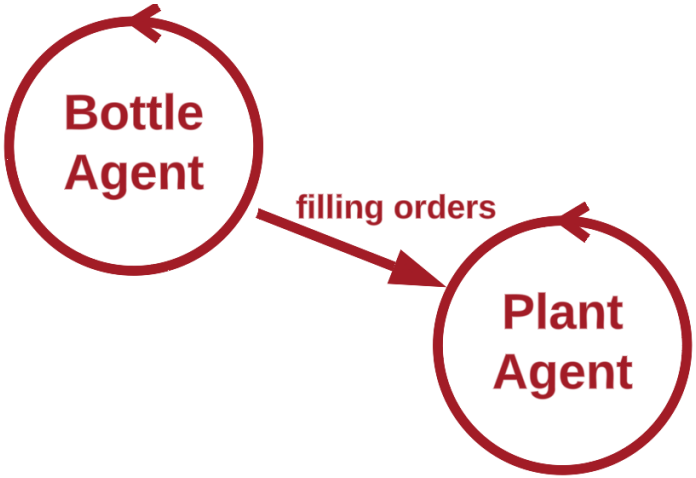
Next in Line...

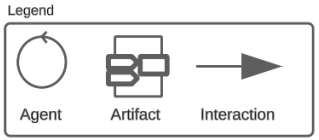
Case Study: Laboratory Plant *myJoghurt*

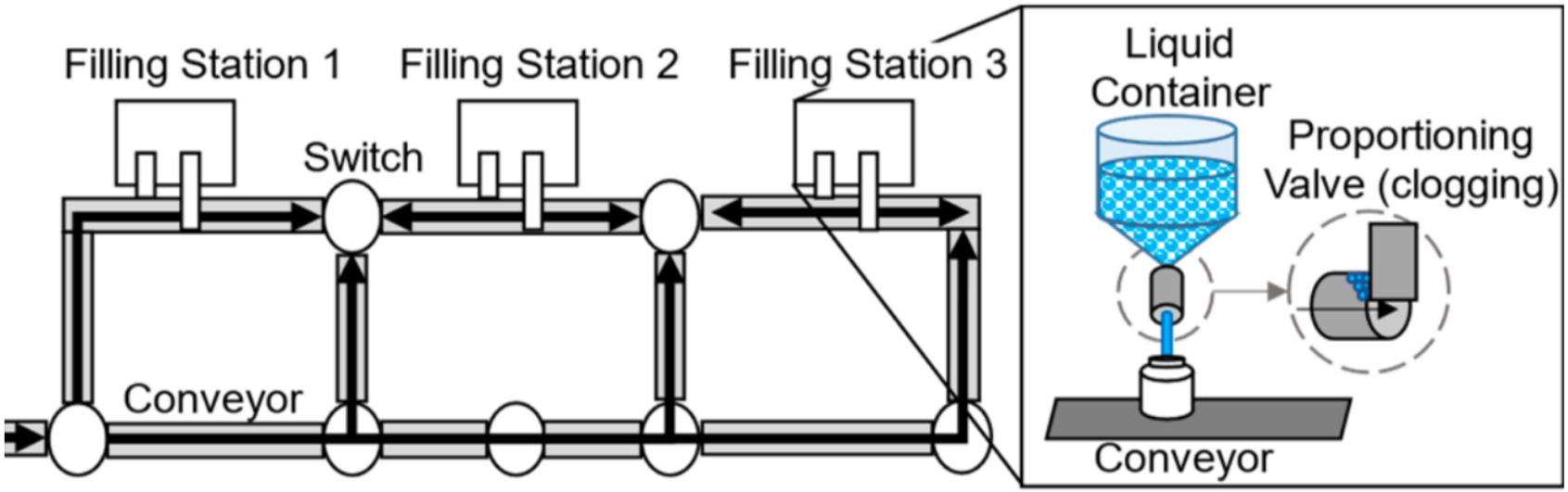




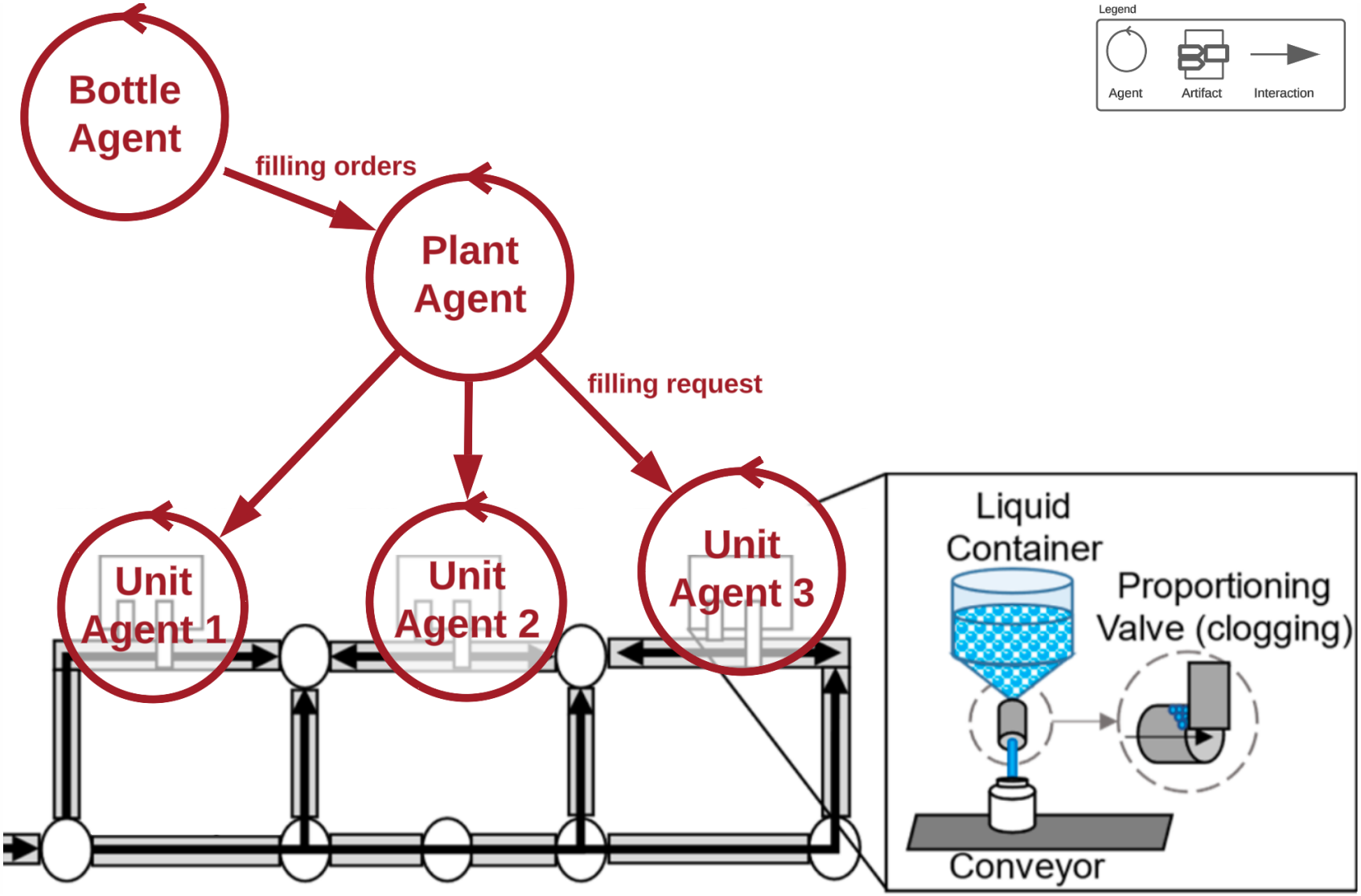
Case Study: A MAOP Design & Programming



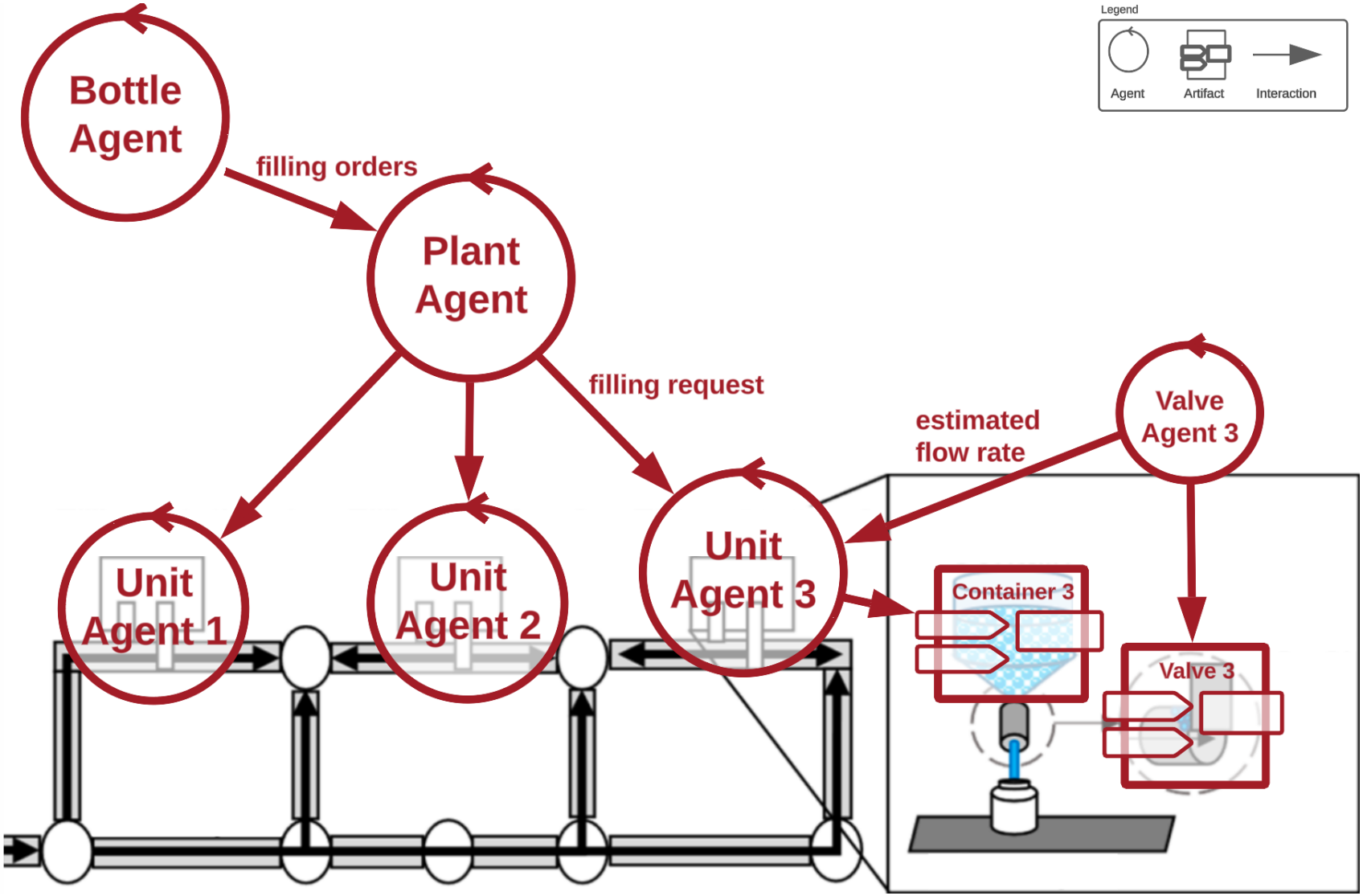




Case Study: A MAOP Design & Programming



Case Study: A MAOP Design & Programming



Problems and Objective

Problems:

- the system is *dynamic* and complex (e.g., duration and degree of the clogging are *non-deterministic*, and *hardly measurable*)
- lots of manual interventions are required in a *hard-coded* solution

Problems and Objective

Problems:

- the system is *dynamic* and complex (e.g., duration and degree of the clogging are *non-deterministic*, and *hardly measurable*)
- lots of manual interventions are required in a *hard-coded* solution

The objective is:

- to allow flexibility and **adaptation** in decision-making by autonomous agents
- while monitoring and **controlling** their behavior at runtime

Problems and Objective

Problems:

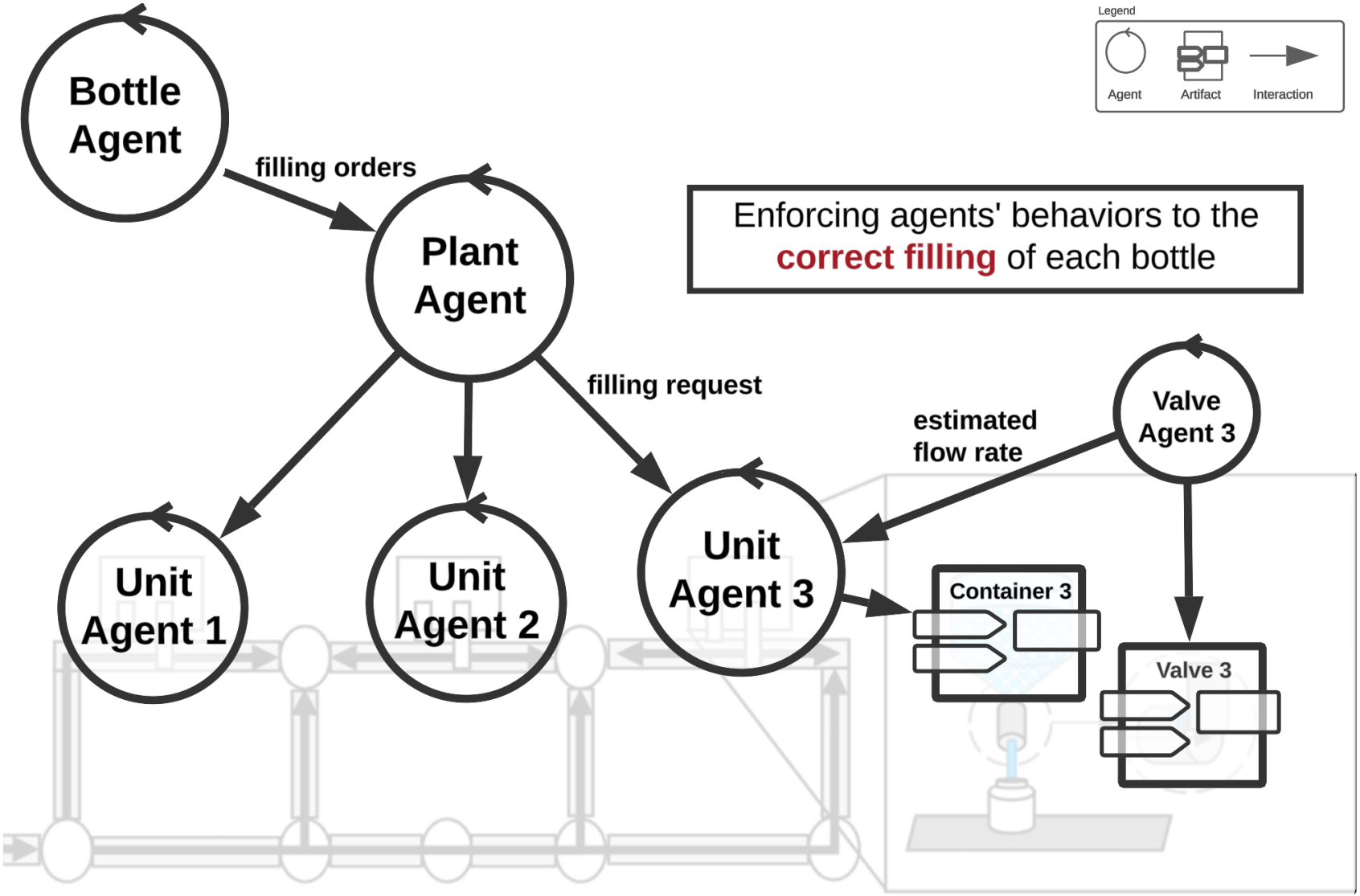
- the system is *dynamic* and complex (e.g., duration and degree of the clogging are *non-deterministic*, and *hardly measurable*)
- lots of manual interventions are required in a *hard-coded* solution

The objective is:

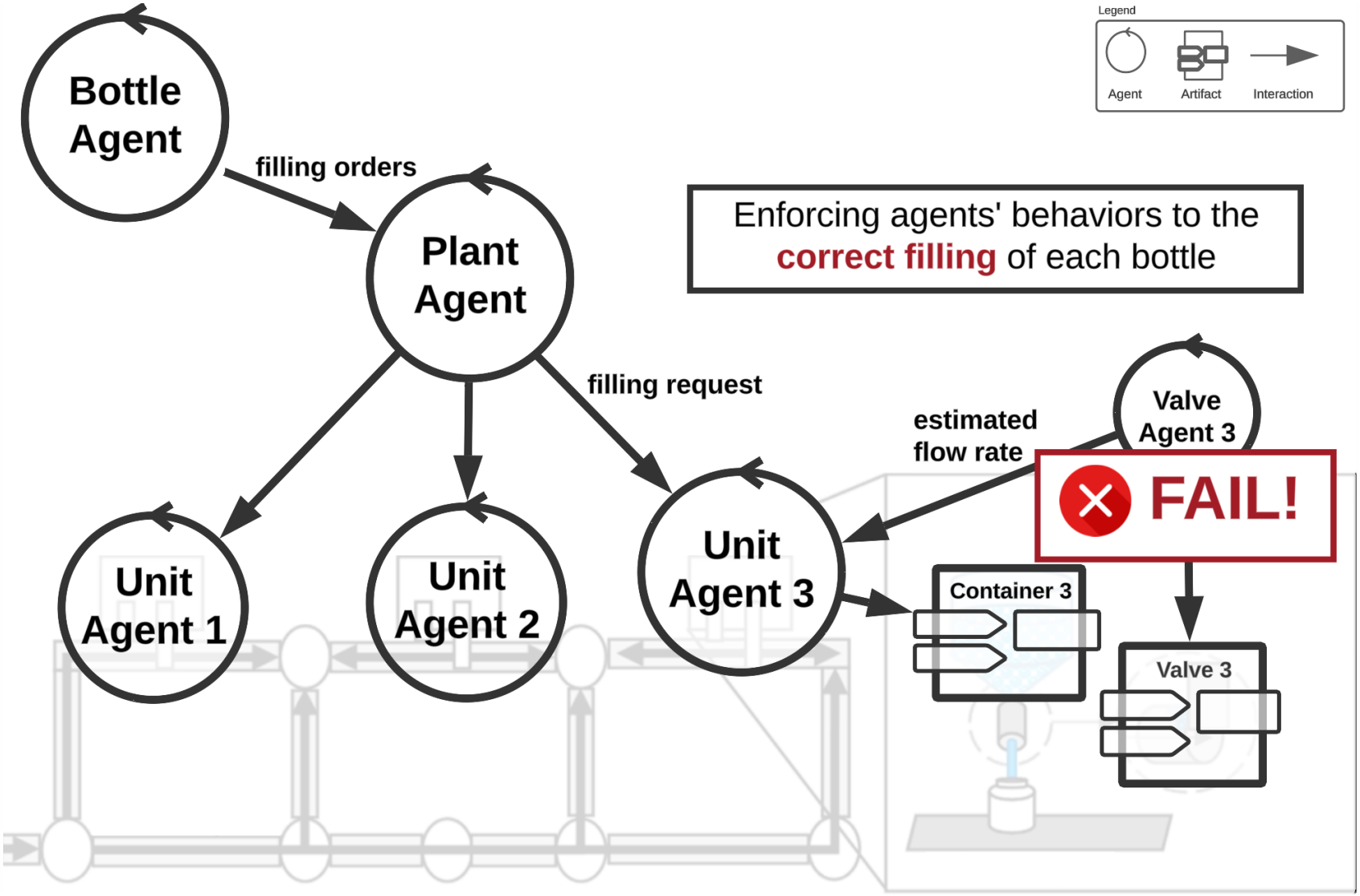
- to allow flexibility and **adaptation** in decision-making by autonomous agents
- while monitoring and **controlling** their behavior at runtime

→ A solution is to have **self-regulation**!

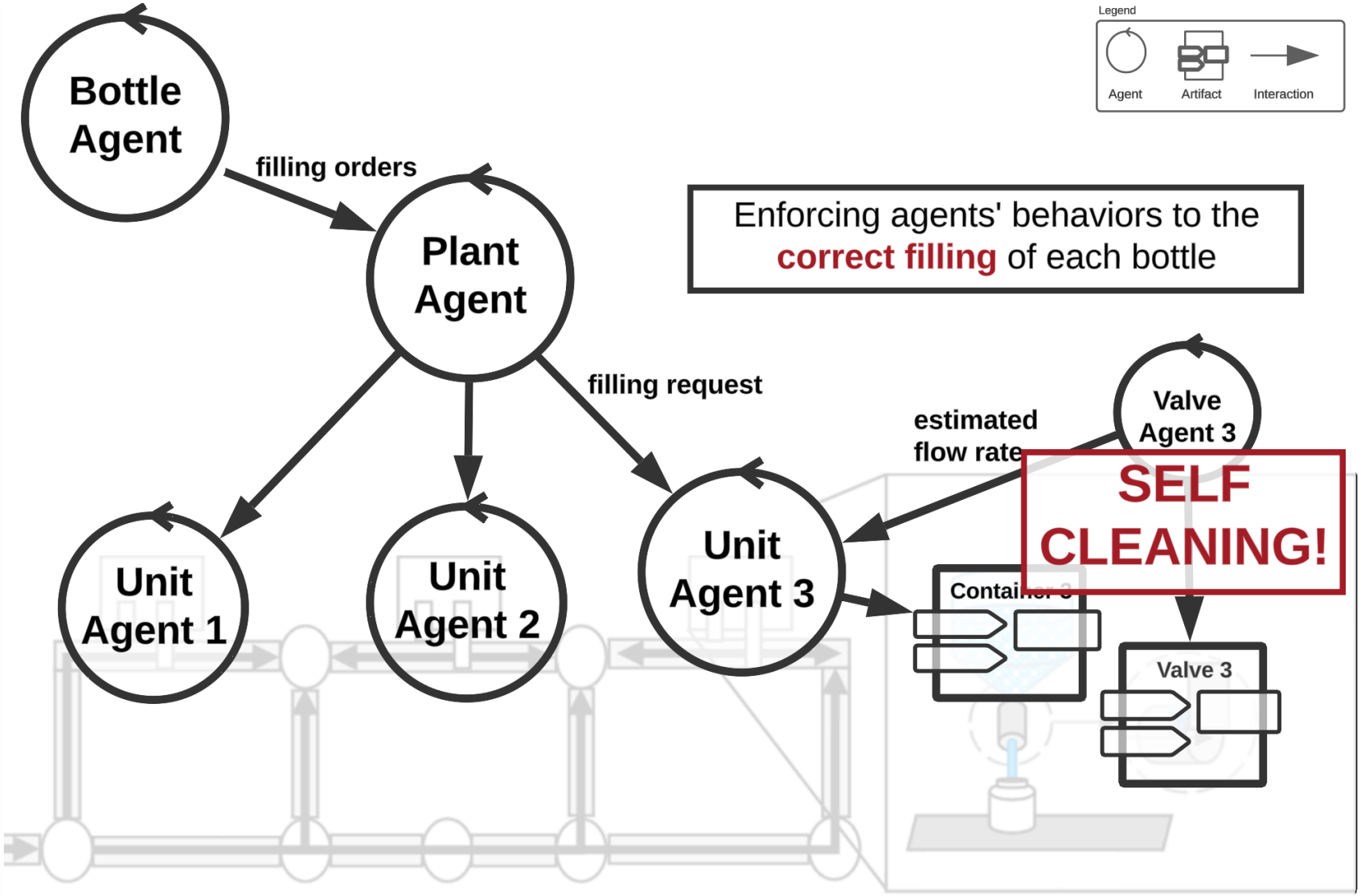
Case Study: How can we regulate the system?



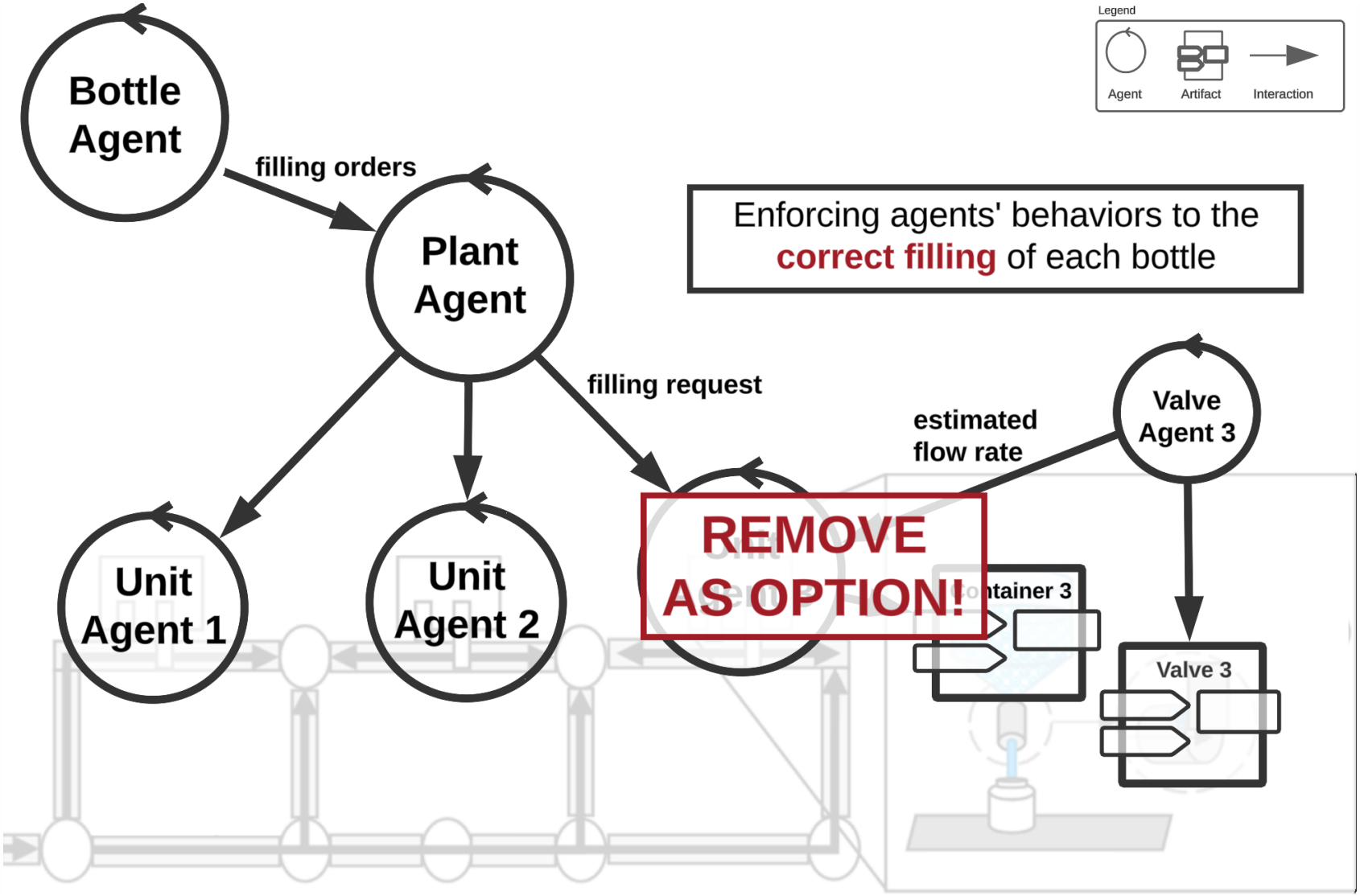
Case Study: How can we regulate the system?



Case Study: How can we regulate the system?



Case Study: How can we regulate the system?



Next in Line...

Research Questions

Our objective is to maintain a balance between agents' autonomy and system regulation.

Research Questions

1. How to express agent's expected behaviors and enforced behaviors?
2. How to enforce agents' expected behaviors?

State of the Art

1. How to express agent's expected behaviors and enforced behaviors?

Language	Expected behaviors	Enforced behaviors
NPL, 2011	**obligations, permissions, prohibitions**	-
NoA, 2002	obligations, permissions, prohibitions	-
N-2APL, 2012	obligations and prohibitions	**sanction**

Extend NPL to NPL(s) with sanctions as a first-class abstraction

State of the Art

2. How to enforce agents' expected behaviors?

Agent Architecture	Enforcement Mechanism
López y López et al., 2006	use secondary norms
n-BDI, 2014	always sanctions
AORTA, 2015	trigger another norm or plan
Normative MDP, 2010	inflict a cost for the violation
EMIL-I-A, 2007	**adaptive sanction**

Embed NPL(s) engine into a BDI normative agent architecture to enable agents to enforce their or the other agents' behavior

Proposed Approach

- Extend NPL to NPL(s) with sanctions as a first-class abstraction
- Embed NPL(s) engine into a BDI normative agent architecture to enable agents to enforce their or the other agents' behavior

Next in Line...

NPL

```
norm <id>
   : <when>
  -> obligation(<who>, <while>, <what>, <deadline>).
```

Example

```
norm n2
    : level(V,X,L) & .my_name(U)
    -> obligation(U, n2, update_factors(V,X,L),
        deviation_factor(X, "negative", _)).
```

NPL(s) - Sanction Rule

```
norm <id>
   : <when>
  -> obligation(<who>, <while>, <what>, <deadline>)
     [if fulfilled: <sanction-rule>*]
     [if unfulfilled: <sanction-rule>*]
     [if inactive: <sanction-rule>*]
  .
```

Example

```
norm n2
    : level(V,X,L) & .my_name(U)
    -> obligation(U, n2, update_factors(V,X,L),
        deviation_factor(X, "negative", _))
    if unfulfilled: s1(V,X), s2(V,X).
```

NPL(s) - Sanction

```
sanction-rule <id>(<args>)
    : <condition>
   -> sanction(<agent>, <description>).
```

Example (Self Cleaning)

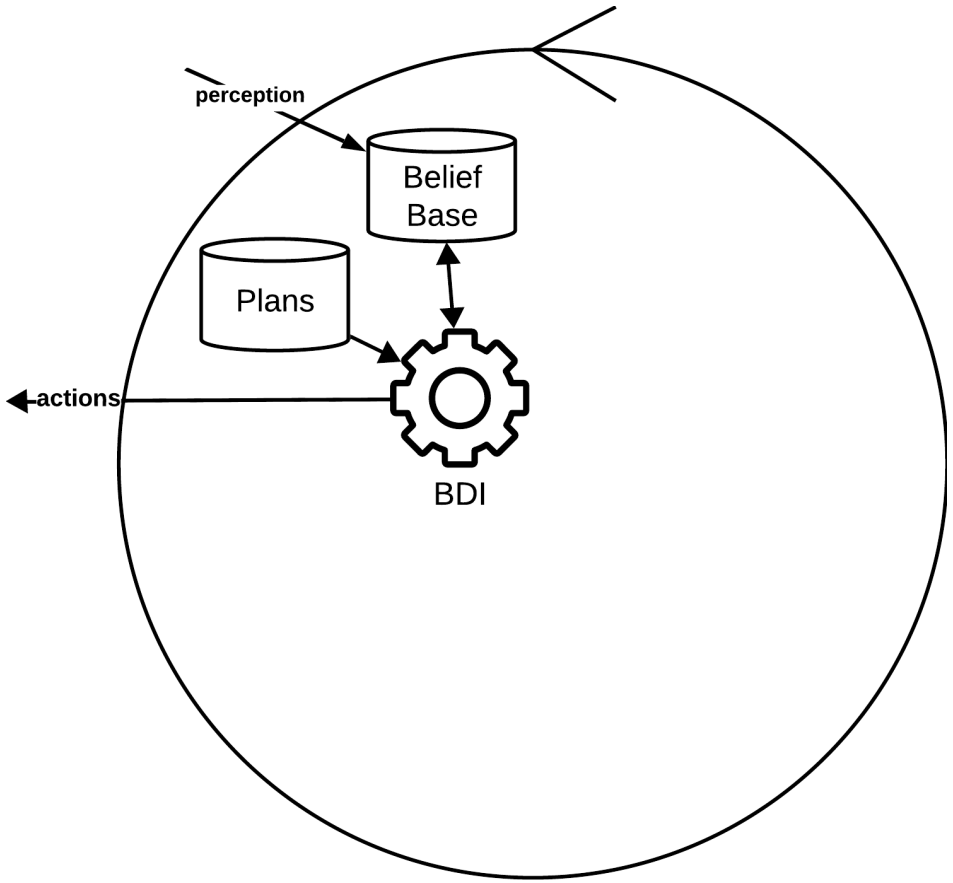
```
sanction-rule s2(V,X)
    : learning_factor(V,X,_,_,_,C) & threshold(_,T) & C>=T
    -> sanction(V, self_cleaning(X)).
```

Id	Sanctioner	Target	Sanction	Condition
S2	unit	valve	Activate the **self-cleaning** procedure	The violation occurs three consecutive times

Next in Line...

Normative Agent Architecture

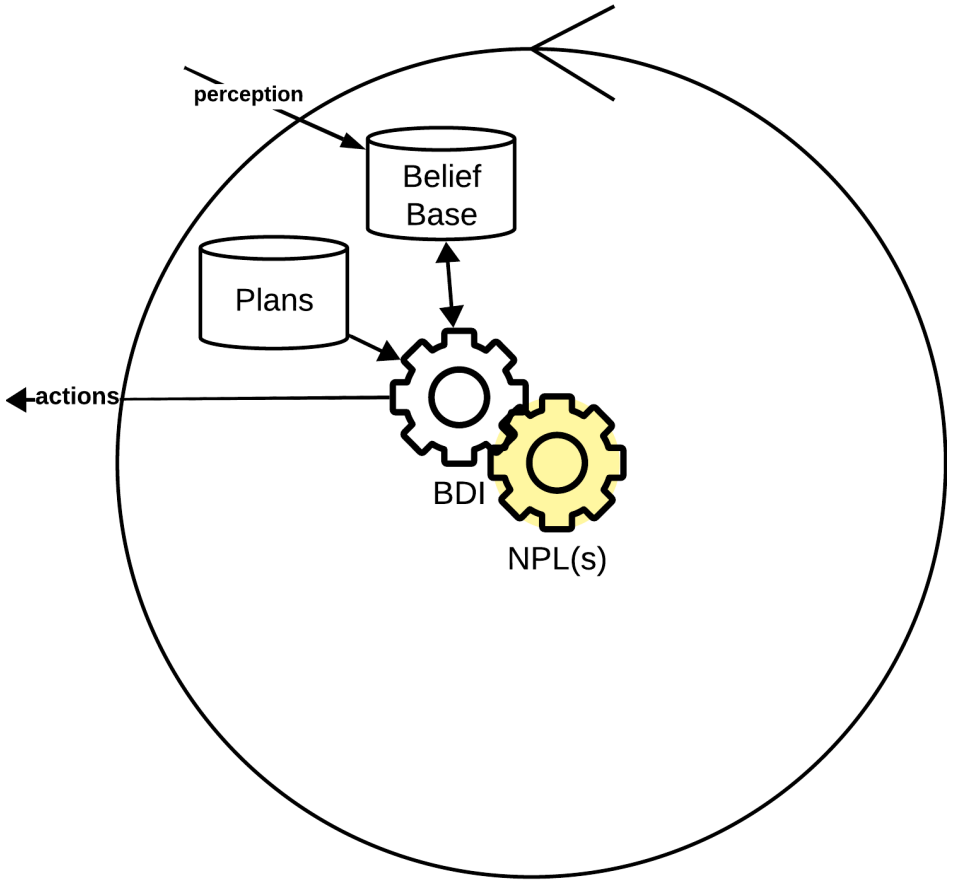
BDI JaCaMo agent architecture



Normative Agent Architecture

We extend the BDI JaCaMo agent architecture by integrating:

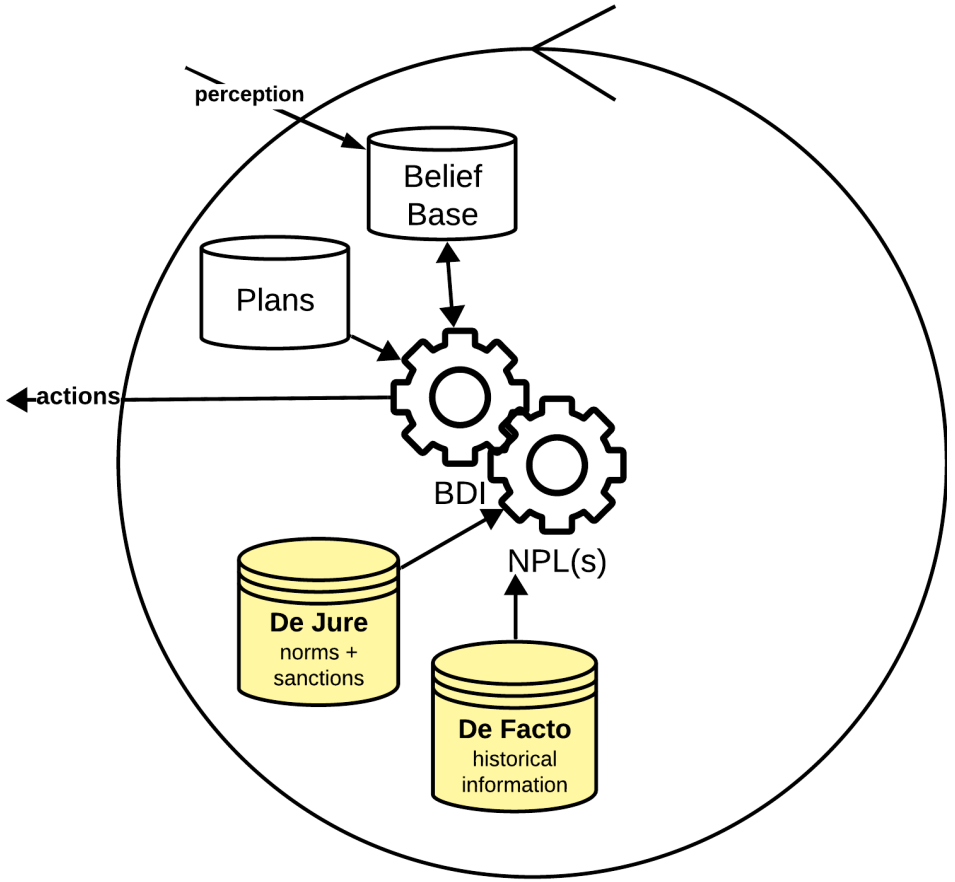
- **NPL(s) Engine**



Normative Agent Architecture

We extend the BDI JaCaMo agent architecture by integrating:

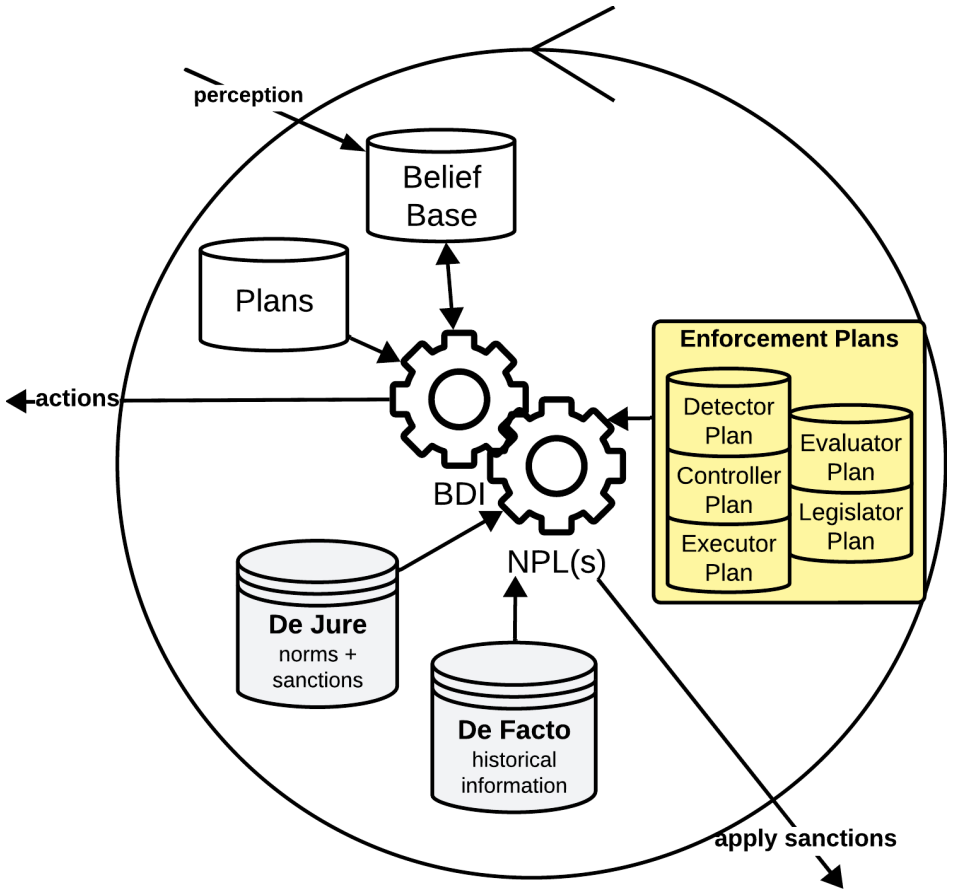
- **NPL(s) Engine**
- **De Jure**
- **De Facto**

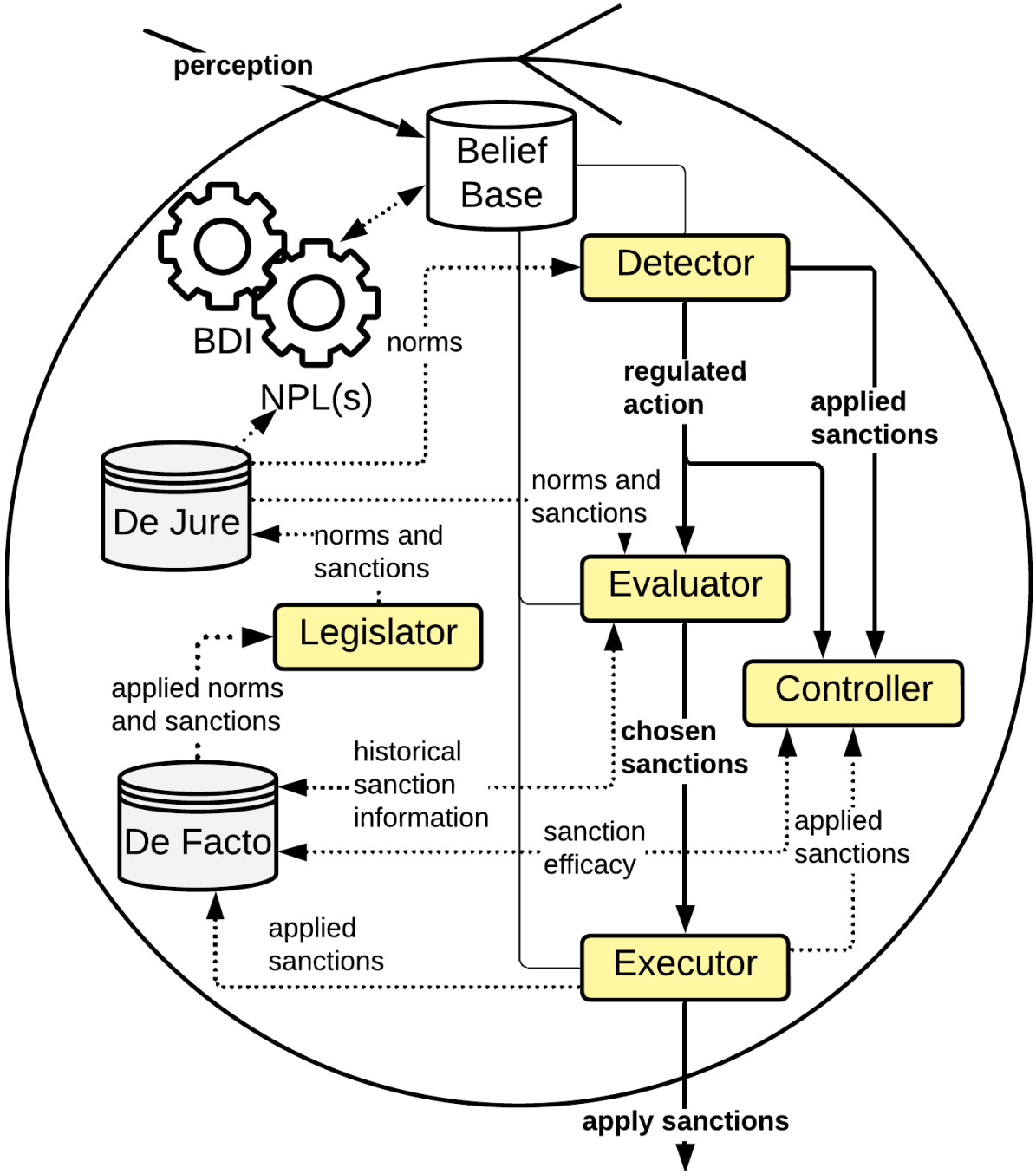


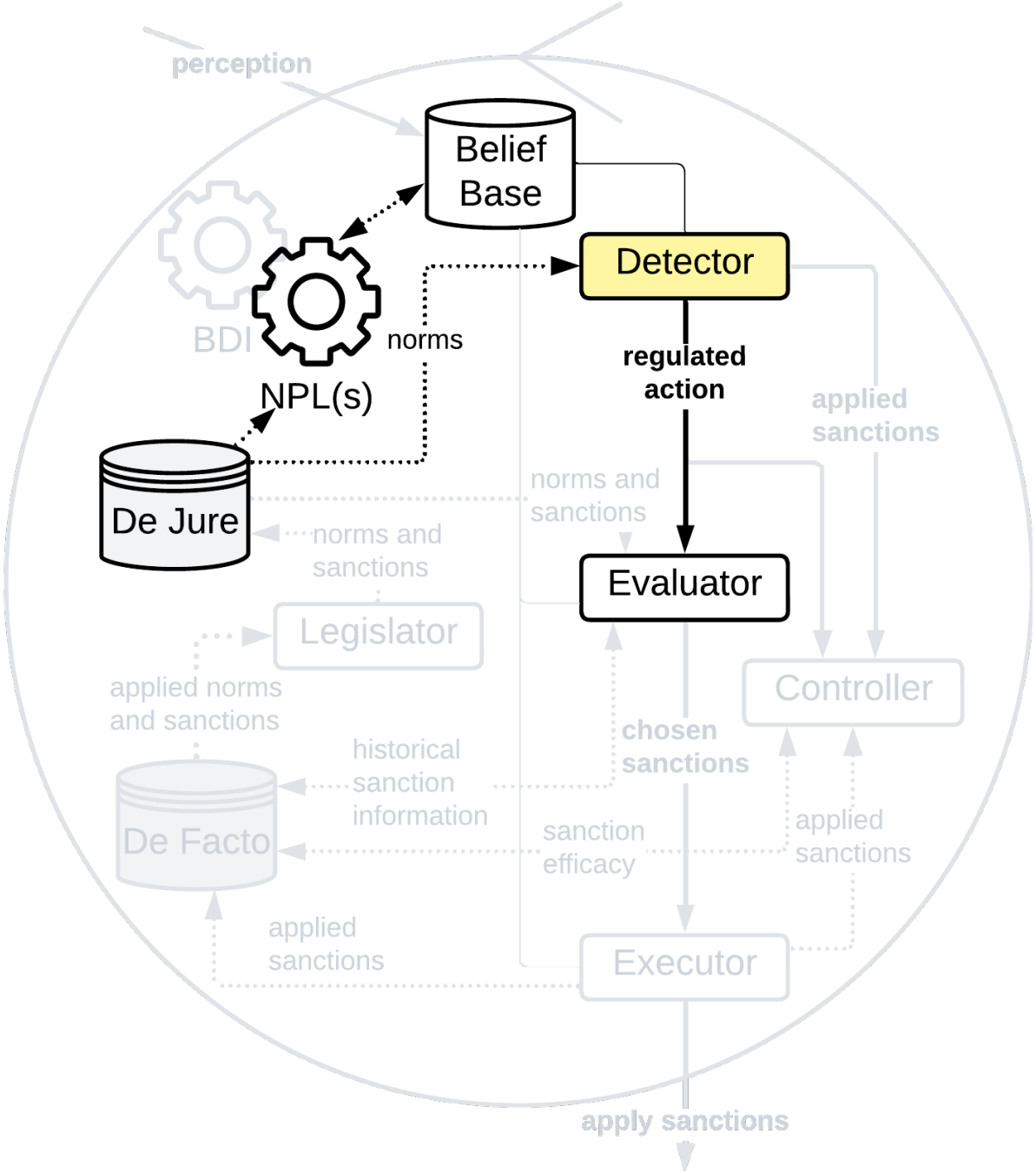
Normative Agent Architecture

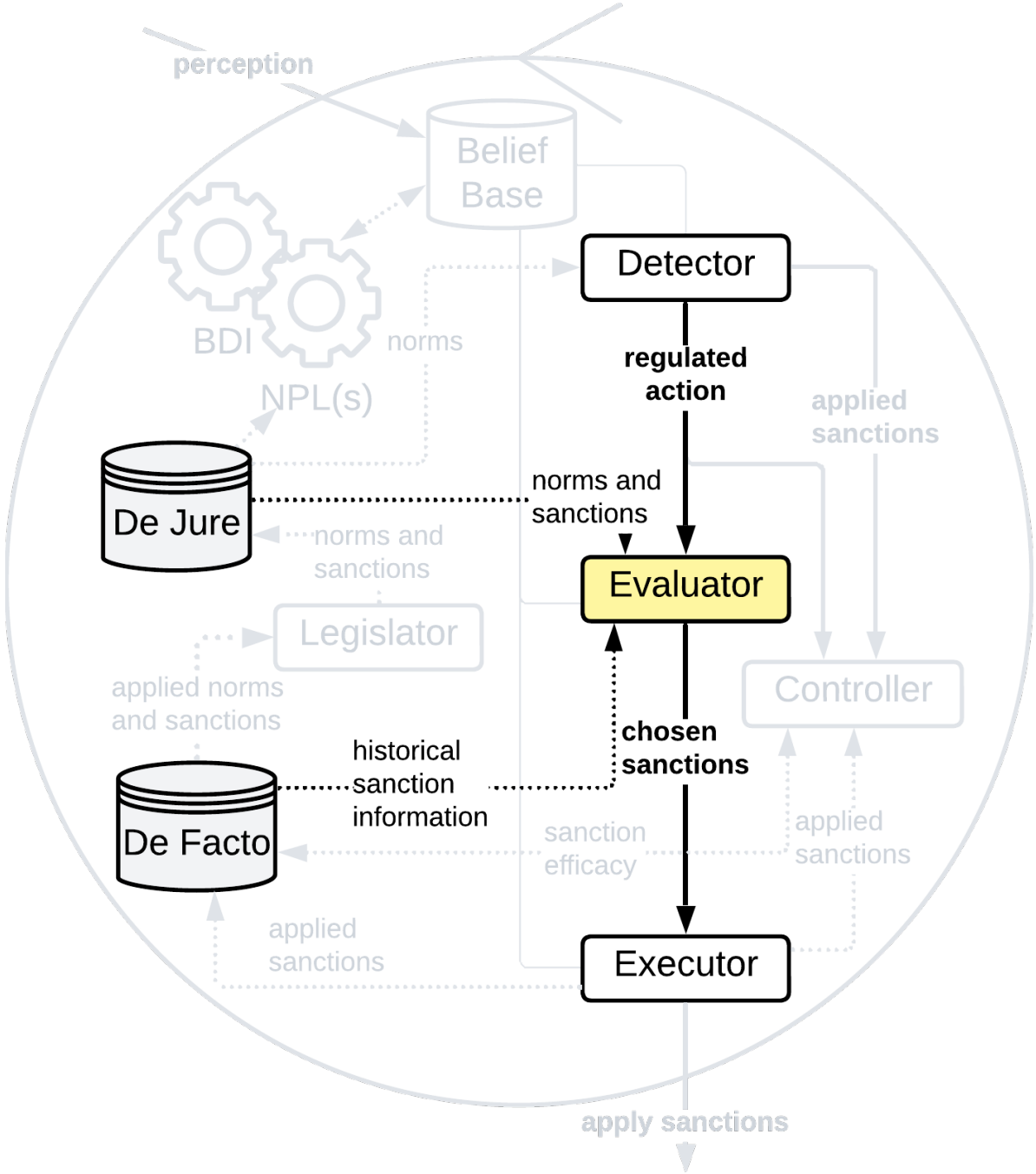
We extend the BDI JaCaMo agent architecture by integrating:

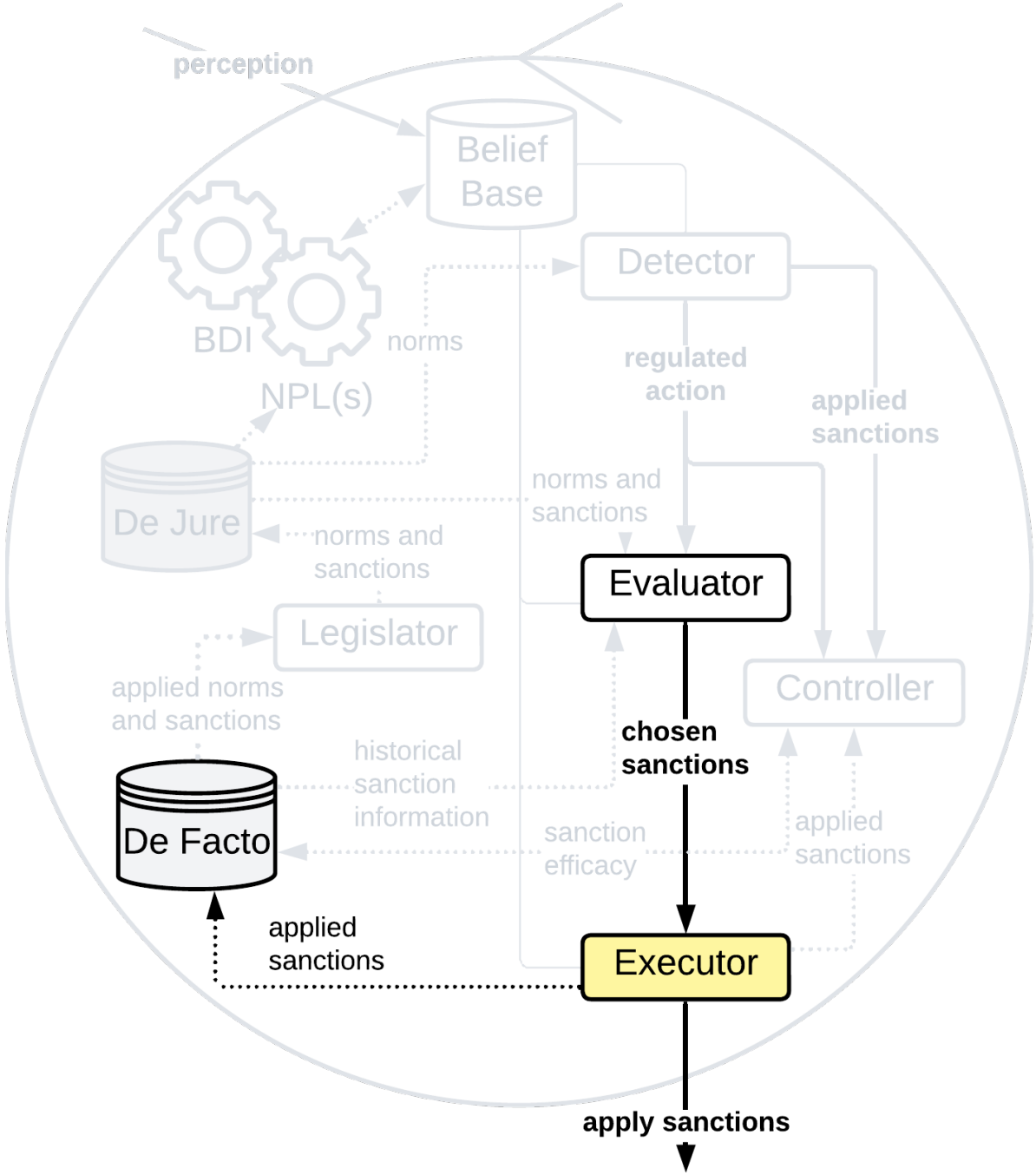
- **NPL(s) Engine**
- **De Jure**
- **De Facto**
- **Enforcement Plans**:
  - Detector
  - Evaluator
  - Executor
  - Controller
  - Legislator

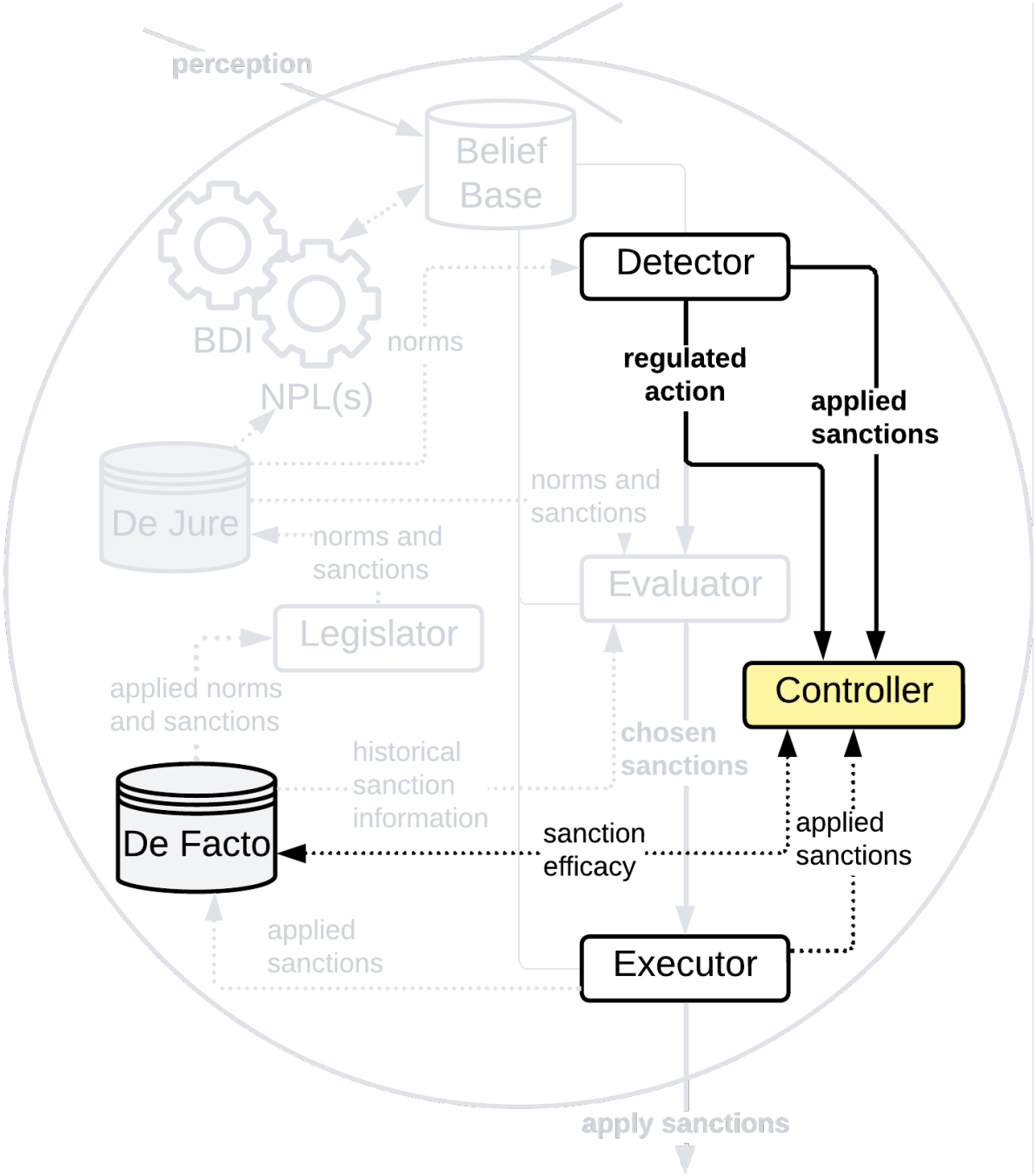


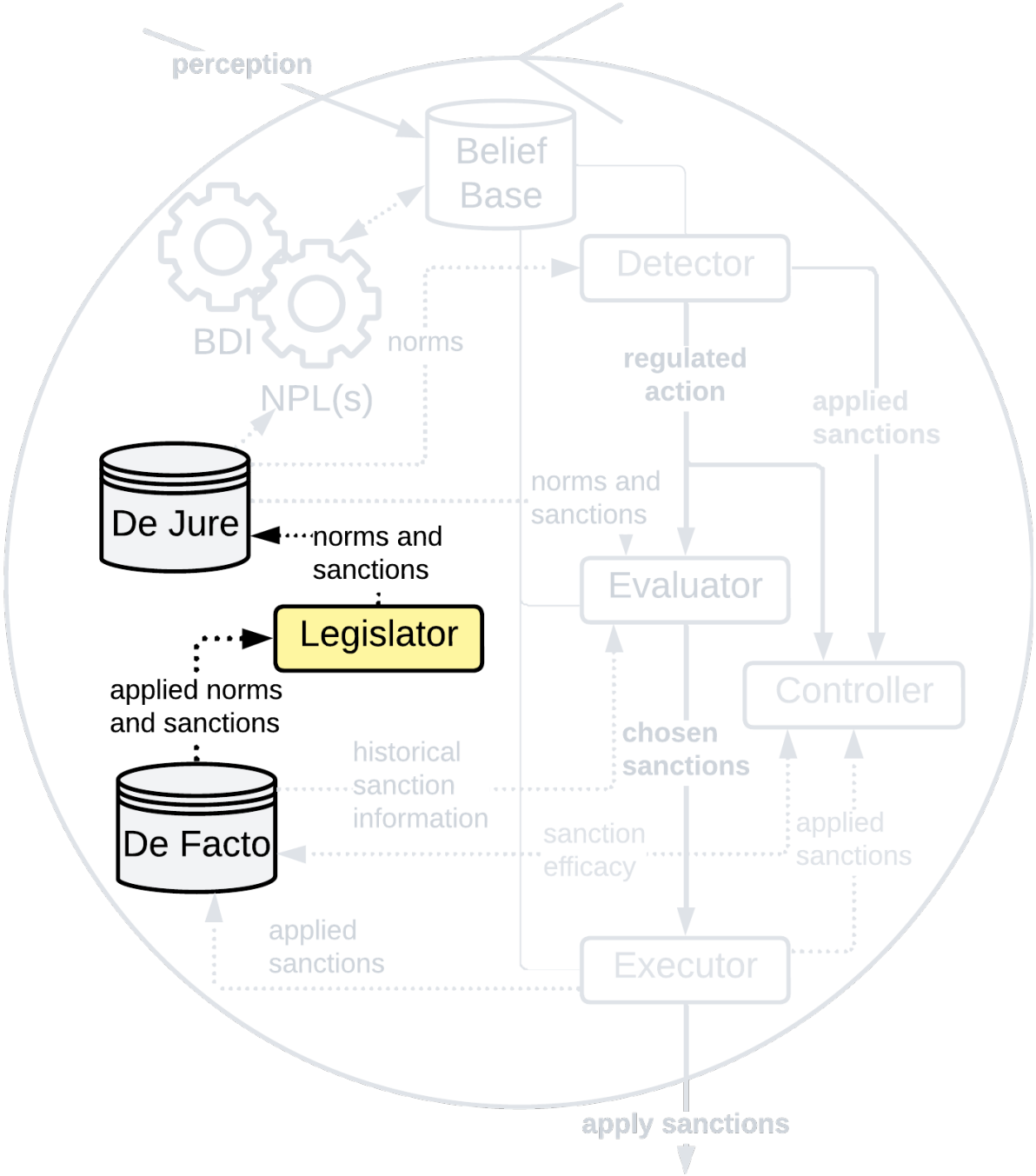

perception
Belief Base
Detector
BDI
NPL(s)
norms
regulated action
applied sanctions
De Jure
norms and sanctions
norms and sanctions
Evaluator
Legislator
Controller
applied norms and sanctions
chosen sanctions
historical sanction information
De Facto
sanction efficacy
applied sanctions
applied sanctions
Executor
apply sanctions

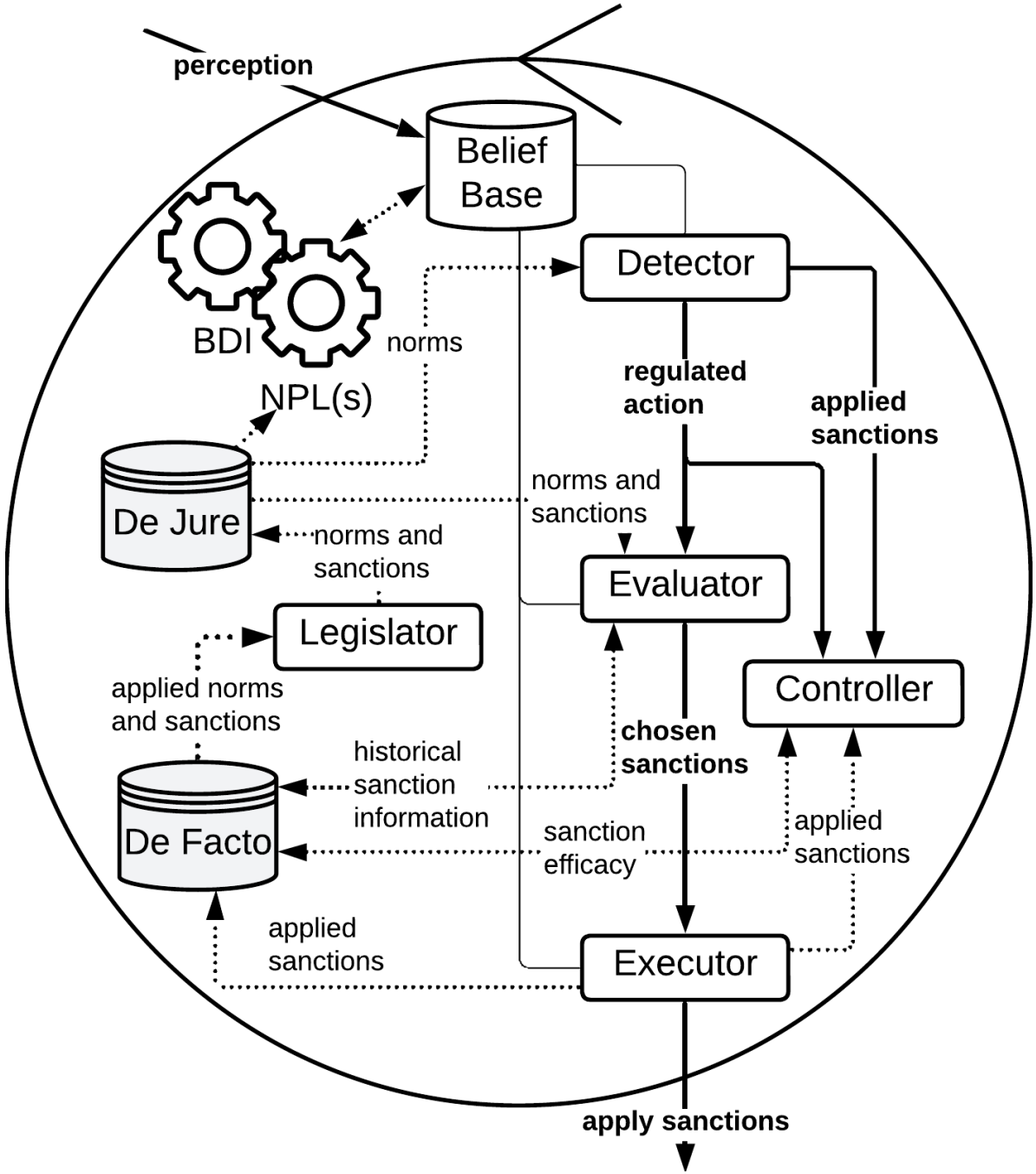

perception
Belief Base
Detector
BDI
norms
NPL(s)
regulated action
applied sanctions
De Jure
norms and sanctions
norms and sanctions
Evaluator
Legislator
Controller
applied norms and sanctions
chosen sanctions
historical sanction information
De Facto
sanction efficacy
applied sanctions
applied sanctions
Executor
apply sanctions


perception
Belief Base
Detector
BDI
NPL(s)
norms
regulated action
applied sanctions
De Jure
norms and sanctions
norms and sanctions
Evaluator
Legislator
applied norms and sanctions
Controller
chosen sanctions
historical sanction information
De Facto
sanction efficacy
applied sanctions
applied sanctions
Executor
apply sanctions


perception
Belief Base
Detector
BDI
NPL(s)
norms
regulated action
applied sanctions
De Jure
norms and sanctions
norms and sanctions
Evaluator
Legislator
Controller
applied norms and sanctions
chosen sanctions
historical sanction information
De Facto
sanction efficacy
applied sanctions
applied sanctions
Executor
apply sanctions


perception
Belief Base
Detector
BDI
NPL(s)
norms
regulated action
applied sanctions
norms and sanctions
De Jure
norms and sanctions
Evaluator
Legislator
Controller
applied norms and sanctions
chosen sanctions
historical sanction information
De Facto
sanction efficacy
applied sanctions
applied sanctions
Executor
apply sanctions


perception
Belief Base
Detector
BDI
NPL(s)
norms
regulated action
applied sanctions
norms and sanctions
De Jure
norms and sanctions
Evaluator
Legislator
Controller
applied norms and sanctions
chosen sanctions
historical sanction information
De Facto
sanction efficacy
applied sanctions
applied sanctions
Executor
apply sanctions

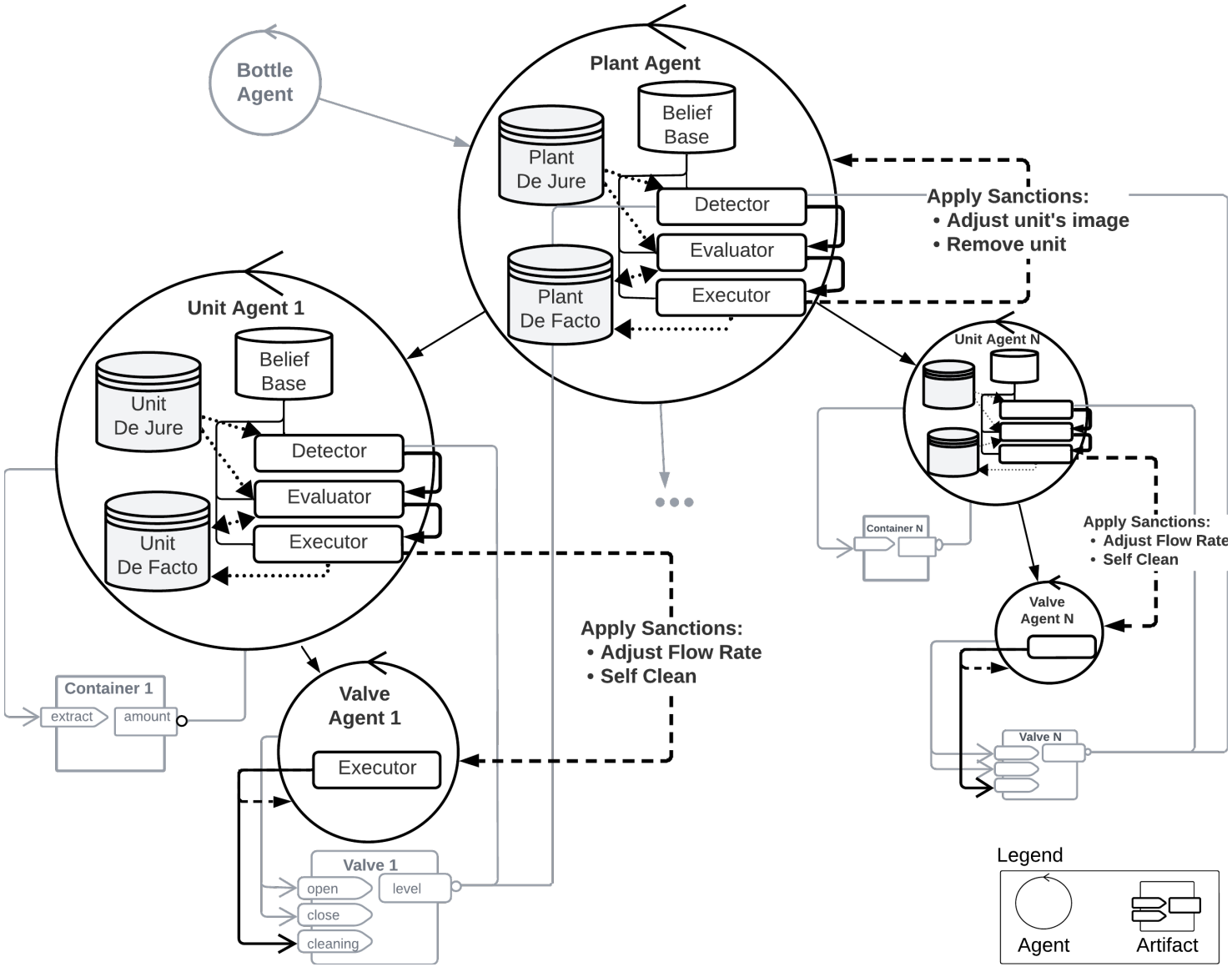

perception
Belief Base
Detector
BDI
NPL(s)
norms
regulated action
applied sanctions
De Jure
norms and sanctions
norms and sanctions
Evaluator
Legislator
applied norms and sanctions
Controller
chosen sanctions
historical sanction information
De Facto
sanction efficacy
applied sanctions
applied sanctions
Executor
apply sanctions

myJoghurt Case Study: Norm and Sanctions

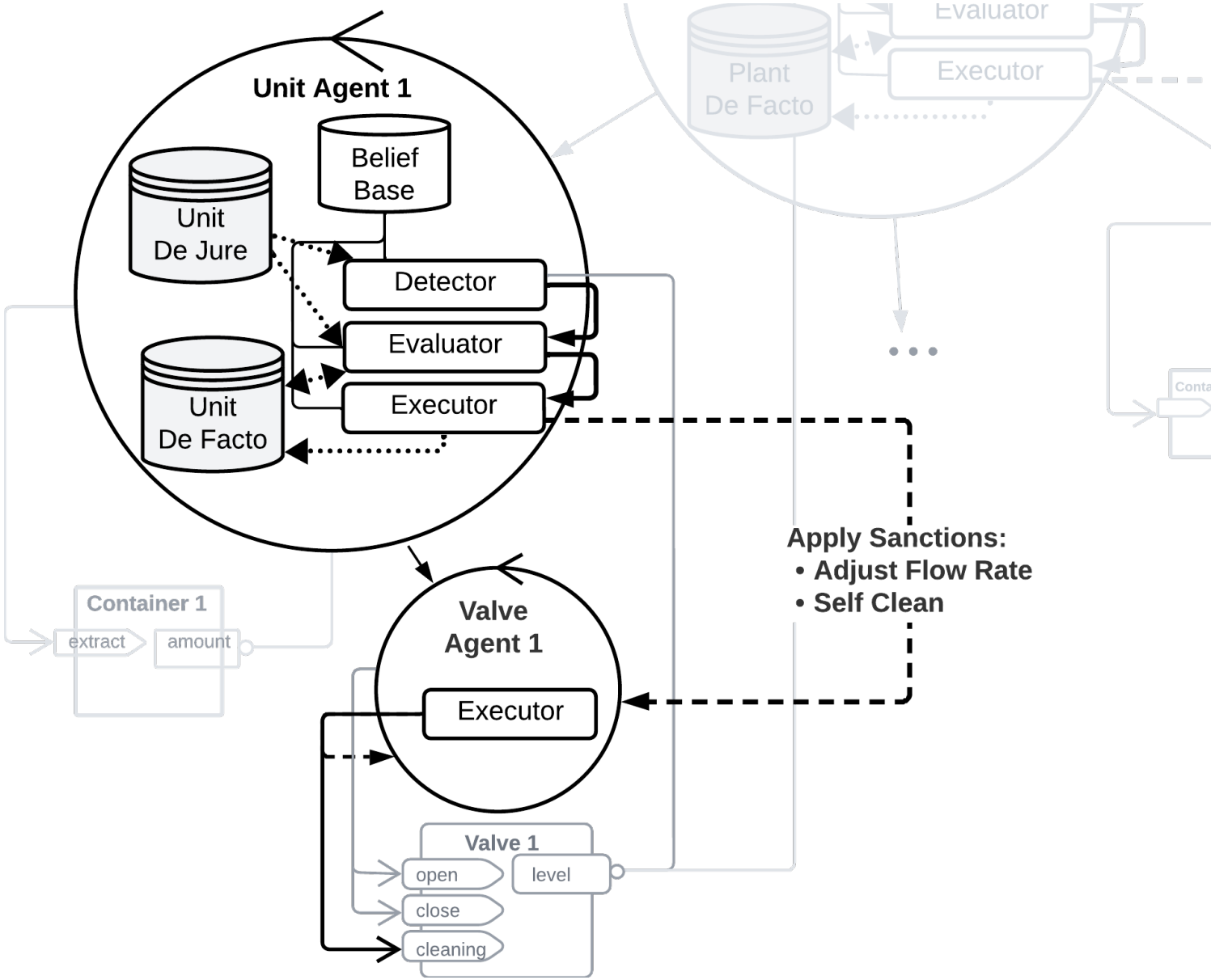
The norms refer to the correct filling of each bottle.

Sanctioner	**Target**	**Sanction**	**Pre-condition**
unit	valve	**Adjust** the estimated **flow rate**	The image is below a threshold
unit	valve	Activate the **self-cleaning** procedure	The violation occurs three consecutive times
plant	unit	**Adjust** the unit agent's **image**	The image is below a threshold
plant	unit	**Remove as an option** for subsequent filling orders	The violation occurs five consecutive times

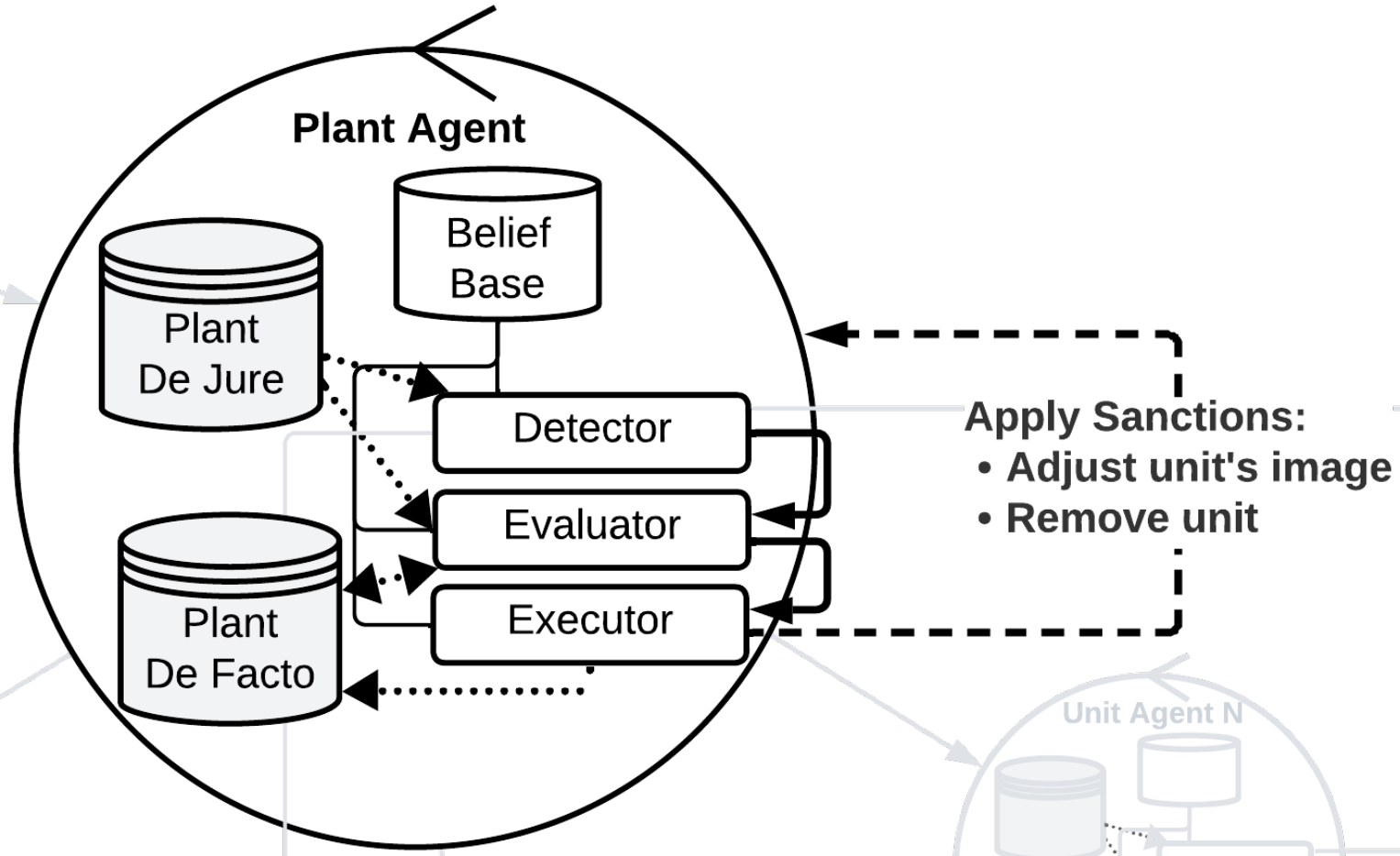
myJoghurt Case Study: MAS Architecture



myJoghurt Case Study: MAS Architecture



myJoghurt Case Study: MAS Architecture



Next in Line...

Conclusions

1. **How to express agent's expected behaviors and enforced behaviors?**

   NPL(s) with the explicit representation of norms and sanctions

2. **How to enforce agents' expected behaviors?**

   Normative agent architecture with a comprehensible and flexible module on norm enforcement and sanctions

$\rightarrow$ Illustrate in an *industrial case study*

Future Work

- Consider the distinction between types of obligations and adapt the sanctioning efficacy
- Connect the sanctioning process directly to the environment with the approach developed in SAI [6]
- Explore accountability [3] and explainability [14] of the normative functioning

Thank you for your attention!

For further information:

Elena Yan, Luis G. Nardin, Jomi F. Hübner, and Olivier Boissier.
An Agent-Centric Perspective on Norm Enforcement and Sanctions.
International Workshop on Coordination, Organizations, Institutions, Norms and Ethics for Governance of Multi-Agent Systems, May 2024, Auckland, New Zealand.
`https://arxiv.org/abs/2403.15128`.

LABORATOIRE D'INFORMATIQUE,
DE MODÉLISATION ET D'OPTIMISATION DES SYSTÈMES

An Agent-Centric Perspective on Norm Enforcement and Sanctions

Elena Yan* Luis G. Nardin* Jomi F. Hübner** Olivier Boissier*

*MINES Saint-Etienne, UMR 6158 LIMOS, France
{elena.yan, luisgustavo.nardin, olivier.boissier}@emse.fr

**Federal University of Santa Catarina, Florianópolis, Brazil
jomi.hubner@ufsc.br

WOA 2024, July 9, 2024

References I

[1] Natasha Alechina, Mehdi Dastani, and Brian Logan.
Programming norm-aware agents.
In Vincent Conitzer, Michael Winikoff, Lin Padgham, and Wiebe van der Hoek, editors, *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, volume 2, pages 1057–1064, Valencia, 2012. IFAAMAS.

[2] Giulia Andrighetto, Rosaria Conte, Paolo Turrini, and Mario Paolucci.
Emergence in the loop: Simulating the two way dynamics of norm innovation.
In *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2007.

[3] Matteo Baldoni, Cristina Baroglio, Roberto Micalizio, and Stefano Tedeschi.
Accountability in multi-agent organizations: from conceptual design to agent programming.
Autonomous Agents and Multi-Agent Systems, 37(1), November 2022.

[4] Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang, and Leendert van der Torre.
The boid architecture: Conflicts between beliefs, obligations, intentions and desires.
In *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 9–16, New York, NY, 2001. ACM.

References II

[5] N. Criado, E. Argente, P. Noriega, and V. Botti.
Reasoning about norms under uncertainty in dynamic environments.
International Journal of Approximate Reasoning, 55(9):2049–2070, 2014.

[6] Maiquel de Brito, Jomi F. Hübner, and Olivier Boissier.
Coupling the normative regulation with the constitutive state management in situated artificial institutions.
The Knowledge Engineering Review, 34, 2019.

[7] Moser Silva Fagundes, Holger Billhardt, and Sascha Ossowski.
Normative reasoning with an adaptive self-interested agent model based on markov decision processes.
In Angel Kuri-Morales and Guillermo R. Simari, editors, *Advances in Artificial Intelligence – IBERAMIA 2010*, pages 274–283. Springer, 2010.

[8] Jomi F. Hübner, Olivier Boissier, and Rafael H. Bordini.
A normative programming language for multi-agent organisations.
Annals of Mathematics and Artificial Intelligence, 62(1-2):27–53, 2011.

[9] Andreas S. Jensen.
The AORTA Reasoning Framework – Adding Organizational Reasoning to Agents.
PhD thesis, Danmarks Tekniske Universitet (DTU), 2015.

References III

[10] Martin J. Kollingbaum and Timothy J. Norman.
Supervised interaction: Creating a web of trust for contracting agents in electronic environments.
In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1*, AAMAS '02, page 272–279. ACM, 2002.

[11] Fabiola López y López, Michael Luck, and Mark d'Inverno.
A normative framework for agent-based systems.
Computational & Mathematical Organization Theory, 12:227–250, 2006.

[12] Felipe Meneguzzi and Michael Luck.
Norm-based behaviour modification in bdi agents.
In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, volume 1, pages 177–184, 2009.

[13] Baldoino F dos S Neto, Viviane Torres da Silva, and Carlos José Pereira de Lucena.
Nbdi: An architecture for goal-oriented normative agents.
ICAART 2011, 2011.

References IV

[14] Elena Yan, Samuele Burattini, Jomi Fred Hübner, and Alessandro Ricci.
Towards a multi-level explainability framework for engineering and understanding bdi agent systems.
In Rino Falcone, Cristiano Castelfranchi, Alessandro Sapienza, and Filippo Cantucci, editors, *Proceedings of the 24th Workshop "From Objects to Agents" (WOA 2023)*, volume 3579 of *CEUR Workshop Proceedings*, Rome, 2023. CEUR-WS.org.