Deliberative Normative Agents: Principles and Architecture

Cristiano Castelfranchi,

Frank Dignum,

Catholijn Jonker, Jan Treur

Social coordination of behaviour

Solutions:

- **protocols (like Contract-Net)**
- **joint goal or intention**

Questions:

- **What to do if the protocol fails?**
- **When to deviate from the plan/protocol?**

Principles

- Norms reduce co-ordination costs
- Norms make institutions and organisation possible and stable
- Norms have positive effects with respect to co-operation, fairness, reduction of aggression
- Intelligent violation of norms can be useful

Deliberative Normative Agents

- **Are able to recognize the existence of norms**
- **Can decide to adopt a norm**
- **Can deliberately follow the norm**
- **or violate it in specific cases**
- **Can react to violations of the norm by other agents**

Deliberative Normative Agents

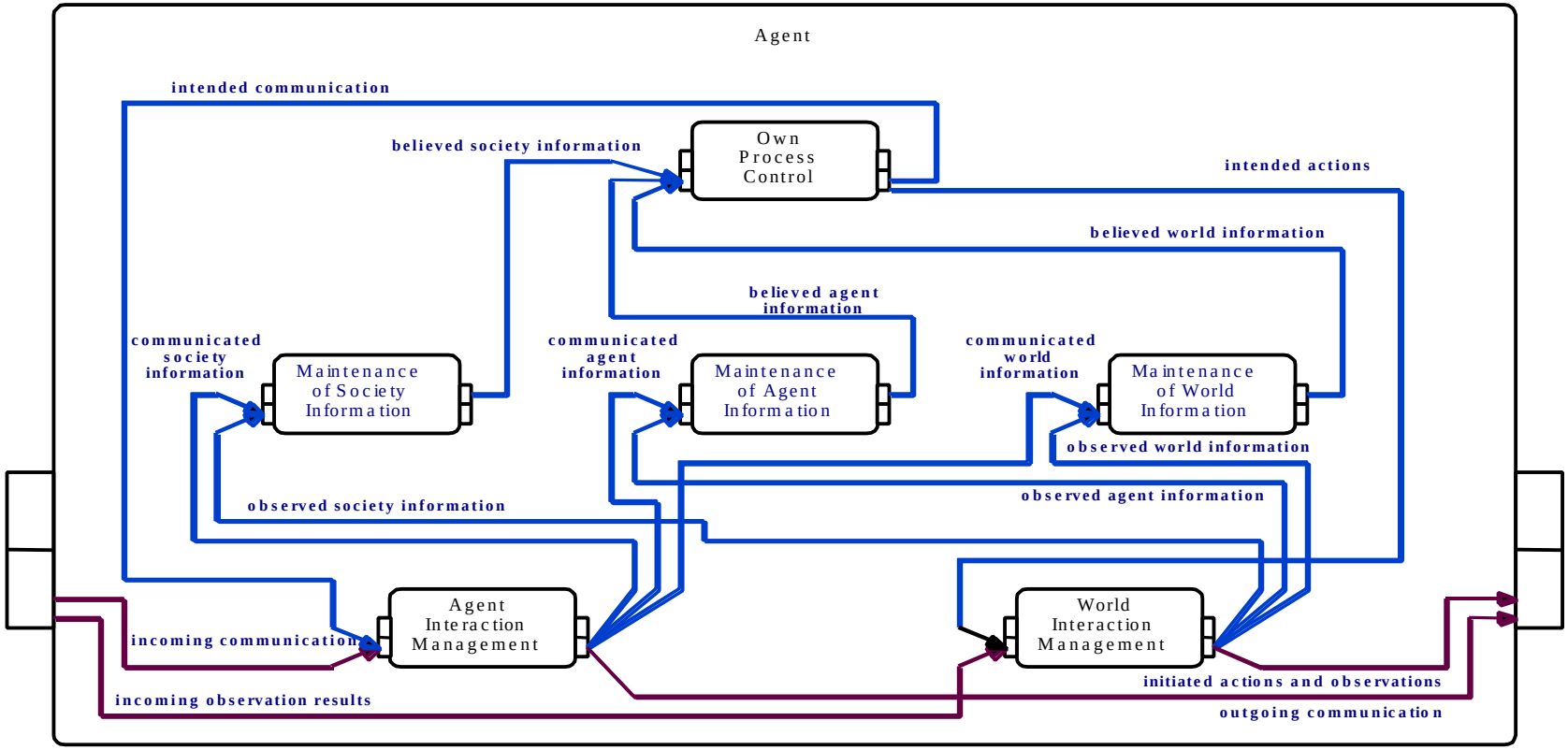
- **Can collectively issue norms**
- **Can reason about norms**
- **Can communicate norms**
- **Can negotiate about norms**

- **Have norms as mental objects**

Consequences for the agent architecture

- how do norms influence the behaviour of the agent?
- the agent should be a cognitive agent:
  it should have some representation of beliefs, goals, intentions, etc.; e.g., the BDI-architecture
- decision to obey the norm is a motivated 'conscious' separate decision
- architecture should include some facility for reasoning about applying the norms and subsequent combination of the result with the goals and actions of the agent

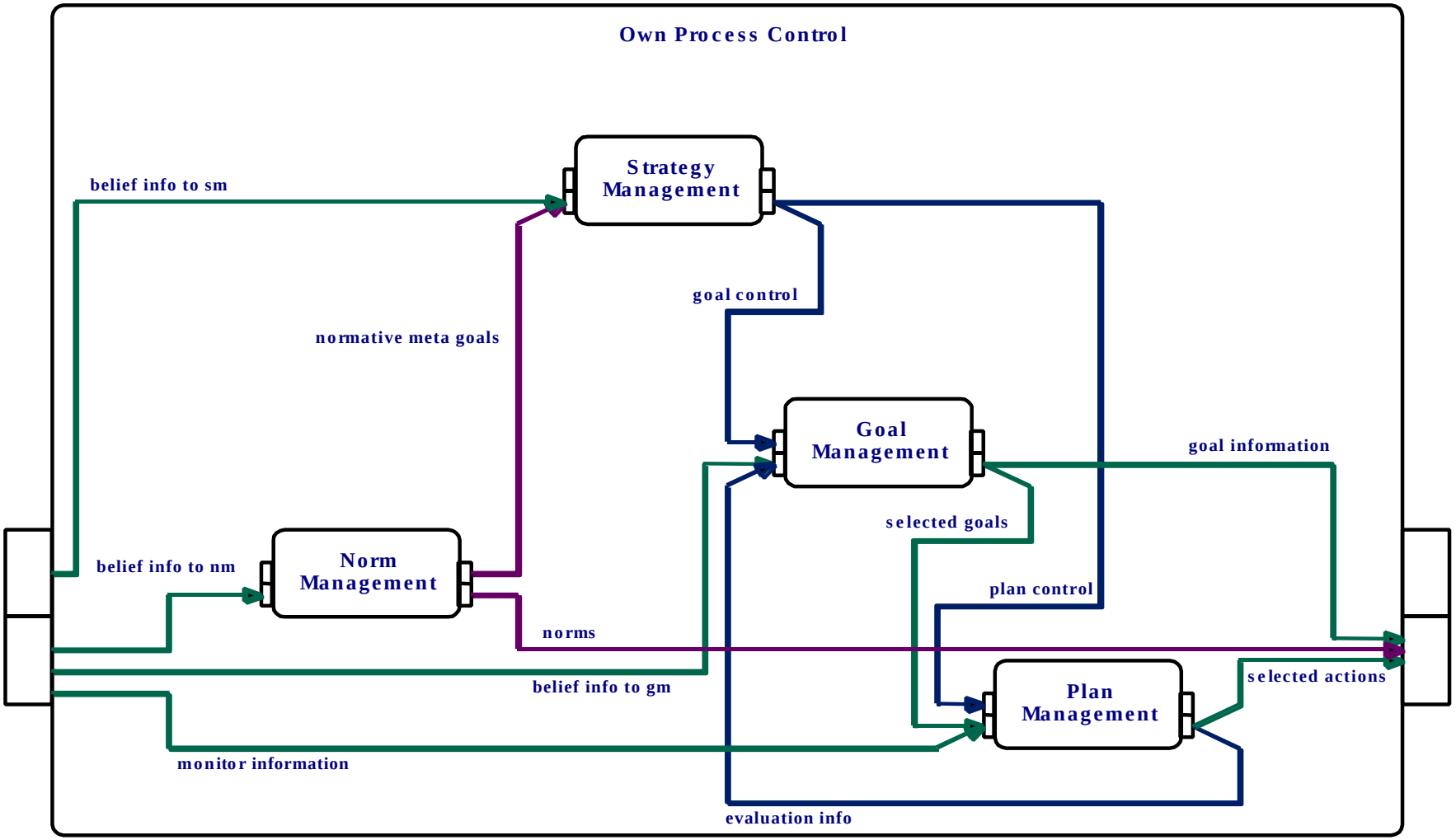
Global architecture



Representation at different meta-levels

- 1. Object level
  - believes like:
    has_norm(society1, you_ought_to_drive_on_the_right)
- 2. Meta-level
  - communicated_by
    (has_norm(self, you_ought_to_drive_on_the_right),
    positive_assertion, agent_B)
- **3.** Meta-meta-level
  - adopted_own_process_goal(you_ought_to_drive_on_the_right)

Own Process Control



Norms and behaviour: 1. Norms and goals

- non-adopted norms:
  - useful for coordination (predict the behaviour of the other agents)
- adopted norms:
  - impact on goal generation; among the possible '*sources of goals*' -> normative goals
  - impact on goal selection by providing criteria about how to select among existing goals; e.g., preference criteria.

2. Norms and plans

- Norms focus plan generation
- Norms may focus plan selection
- Norms may focus action selection

E.g. the norm “be kind to colleagues” may lead to a preferred plan to reach a goal within an organisation.

Types of normative agents

- Agent has no knowledge about norms
- Agent knows about norm but is not influenced
  - The norm is build in. E.g. do not cheat
  - The agent thinks the norm is not applicable. E.g. use indicator when turning

Types of normative agents II

- Agent is influenced
  - agent does not adopt the norm but its plans are influenced by risk of violating the norm
  - agent adopts norm
    - agent follows the norm whenever possible
    - agent violates the norm (sometimes)
    - agent violates the norm always if possible

Discussion

- **deliberation about norms provides a flexible form of coordination in societies**
- **to achieve this, a cognitive agent architecture is required with norms as explicit representations**
- **the interplay between norms, goals and plans has to be addressed**
- **an operational architecture for deliberate normative agents has been designed within the distributed DESIRE software environment**
- **future works: social simulation experiments on the basis of this architecture**