Empowering BDI Agents with Generalised Decision-Making

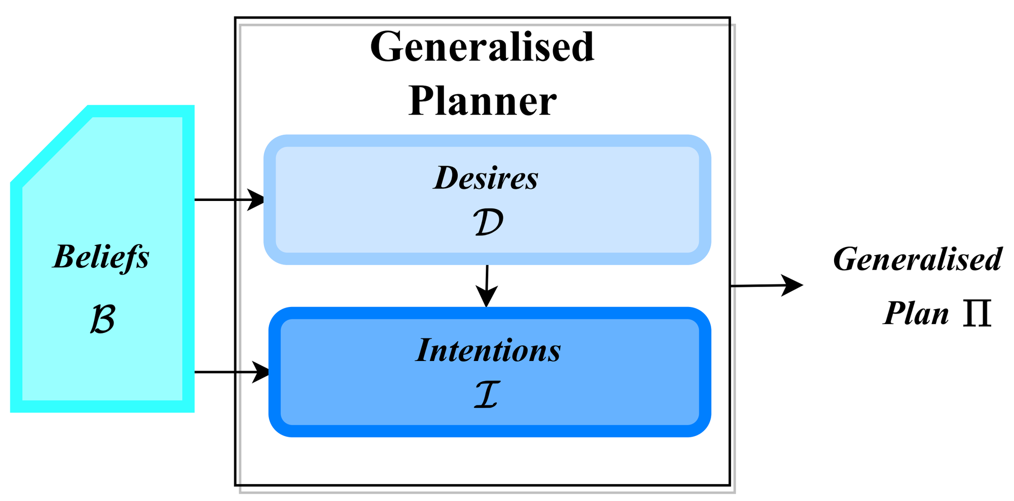
Ramon Fraga Pereira, and Felipe Meneguzzi

University of Manchester (England, UK) – University of Aberdeen (Scotland, UK)

✉ ramon.fragapereira@manchester.ac.uk ✉ felipe.meneguzzi@abdn.ac.uk

Motivation and Goals

- Means-ends reasoning is a critical capability of BDI-style Agents, for which automated planning is a natural class of algorithms
- Glaring disconnect between planning research and agents research
- This Blue Sky paper formalises BDI agents as generalised planners, laying the foundation for future work on planning agents such that:
  - Agents can reason about committing to multiple sets of desires in terms of generalised planning
  - Provides the underpinning for agents to reason about others using a Theory of Mind for Intent and Goal recognition



Generalised Planning

- A variation of the planning problem in which there are multiple initial states and goals $\mathcal{GP} = \langle \mathcal{P}_0, \mathcal{P}_1, ..., \mathcal{P}_N \rangle$
- Solution to a generalised planning problem is analogous to an agent problem

- Provides a natural way to implement means-ends reasoning

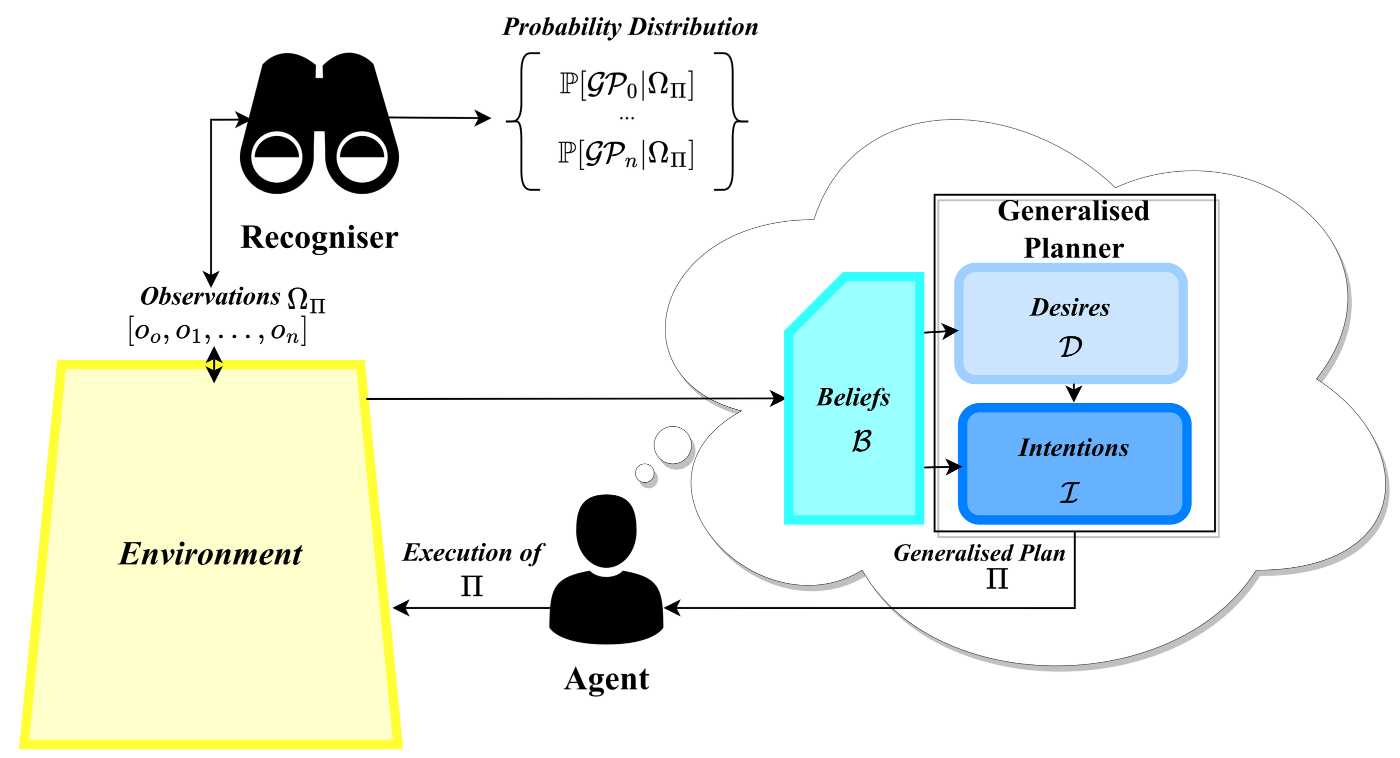
BDI Agents as Generalised Planners

- The BDI architecture has a long tradition in agents research
- Based on the mental attitudes of Beliefs, Desires and Intentions ($\langle \mathcal{B}, \mathcal{D}, \mathcal{I} \rangle$)
- We formalise a BDI architecture with:
  - Exclusively declarative desires $\langle \varphi_i, D_i, \sigma_i \rangle$
  - Means-ends reasoning driven by a generalised planner

procedure REASONINGCYCLE($\mathcal{B}, \mathcal{D}, \mathcal{I}, \Xi$)
  loop
    $\mathcal{B} \leftarrow \mathcal{B} \cup$ SENSE( )
    while $\mathcal{I}$ is not empty **do**
      Pick an intention $\langle\langle \varphi, D \rangle, \pi \rangle \in \mathcal{I}$ s.t. $\mathcal{B} \models \varphi \wedge \neg D$
      ACT(π)
    Find $\{\langle \varphi_1, D_1 \rangle \ldots \langle \varphi_n, D_n \rangle\} \in \mathcal{D}^2$
    s.t. $\exists \Pi$, $\Pi = \mathcal{G}$PLANNER($\{\langle \Xi, \mathcal{B}, D_1 \rangle \ldots \langle \Xi, \mathcal{B}, D_n \rangle\}$)
    $\mathcal{I} \leftarrow \{\langle\langle \varphi_1, D_1 \rangle, \Pi \rangle, \langle\langle \varphi_n, D_n \rangle, \Pi \rangle\}$

Generalised Intent Recognition

- Generalised recognition problem assumes a BDI agent driven by a generalised planner (i.e. drive by multiple concurrent desires): $\langle \mathbb{G}, \Omega_{\Pi} \rangle$, where $\mathbb{G} = \langle \mathcal{GP}_0, \mathcal{GP}_1, ..., \mathcal{GP}_N \rangle$
- Solution to this problem is a probability distribution over desires: $\mathbb{P}(\mathcal{GP} \mid \Omega_{\Pi}) = \eta * \mathbb{P}(\Omega_{\Pi} \mid \mathcal{GP}) * \mathbb{P}(\mathcal{GP})$



Challenges and Opportunities

Promising Approaches

- Our model provides a general *Theory of Mind* for agents reasoning about other agents
- Allows an agent to explicitly reason about the goals of others, and determine their attitude: *cooperative* or *adversarial*

Challenges and Opportunities

- Planning in general is computationally hard, thus, research on desire filters is critical
- Current assumption of knowledge about the goal hypothesis space is a limiting factor
- Need effective ways of determining the *level of rationality* of other agents
- Reasoning cycle needs to expand towards desire failures, reconsideration and replanning
- Hard to predict actions when interleaving intentions