Review of: "Metacognitive Agents for Ethical Decision Support: Conceptual Model and Research Roadmap"

John Páez[1]

1 Universidad Distral "Francisco José de Caldas"

Potential competing interests: No potential competing interests to declare.

Modeling metacognitive processes is an area of interest in Robotics and Artificial Intelligence. Metacognition includes at least two aspects: Knowledge (Agent, Task, Strategies) and Control (Planning, Evaluation, Supervision) of cognitive activities. Metacognition occurs in agents. The agent model proposed in the document is BDI. The agent's wishes are linked to the "Systems" proposed by Kahneman. The proposed conceptual model should be complemented. It is suggested to refine the explanation with aspects such as:

- Methods for assessing the agent's beliefs.
- Agent's wishes selection model. For example, explain if there is competition for the Agent's Goals. It is suggested to explain how the Agent's Desires Module is articulated with the two Kahneman Systems (Fast, Slow).
- Expand the relationship of the agent's resources and abilities with the development of the agent's Intentions.

Finally, the document presents three challenges:

- Model artificial metacognition. (See: The Taxonomy of Metacognition by Pina Tarricone).
- Articulate the BDI proposal with the metacognition model.
- Be located in the "Gray" line of the concept of ethics.

It is suggested to reduce the challenges and focus on the architecture.