The overlooked role of self-agency in artificial systems

Matthew D. Goldberg[1], Justin Brody[2], Timothy C. Clausner[1], and Donald Perlis[1]

[1] University of Maryland, College Park
[2] Goucher College

Many AI systems, including vision and language applications, arguably show insufficient attention to what can broadly be described as concepts of self. Here we constrain the focus to shortcomings that we believe would be addressed by the inclusion of two components of a general conceptualization of self, and would particularly contribute to a system's sense of self-agency: a spatial sense of self, and self-reasoning about its own changing/evolving belief history.

A robotic system needs a representation of itself in space as it performs tasks which call for investigation of its environment. Even very simple-seeming behaviors tend to hide complex self-time-body-space issues. Consider being asked "Is there any milk in the fridge?"—a mere "yes," "no," or "I don't know" seems enough. But to have any broad usefulness this will depend on the agent knowing that it is being addressed; knowing whether it knows about the milk; knowing that if it does not know then it might easily find out (open the fridge and look); coming to know that it is now undertaking to open the fridge, and then seeing that now it has succeeded; knowing that there are several cartons of juice along the front of the top shelf, but no milk is visible; knowing that it cannot see through the juice containers and that there might be milk behind them; knowing that it can move the juice to see behind, and seeing that it is now doing that; seeing that there is no milk and inferring that it now knows the answer to the original question; and so on. Every step above involves reasoning about own-knowledge and own-action, applied to dialog and active vision. Without a general mechanism to represent self-in-space as it performs an evolving sequence of actions, AI is currently stuck with a multitude of single-purpose highly constrained and largely inflexible behaviors, and no understanding that any of those behaviors involve itself or the world around it.

With respect to self-reasoning, specifically in dialog, we offer an example of a system's ability to represent itself as an agent with beliefs evolving over time. We argue that this ability allows for greater capability in the area of meta-reasoning and in particular in meta-dialog. For example, AI agent A is conversing with agent B about someone named John. As the conversation proceeds, agent A infers that B is thinking of a different John [1]. If A has an explicit representation of its beliefs changing over time and is able to retain a direct account of the evolution of those beliefs, then it can make such an inference. The conclusion is enabled by reasoning about self in relation to word meanings.

Related to a system's monitoring of its reasoning, Brody et al. [2] advocated for the utility of a computational process that has a highly developed sense of

internal actor and observer that merge in real time. That is, within a narrow interval of time, there is a process of monitoring and adjusting the agent's own processing. As a step toward a realization of this, which also involves language generation from a robot, Brody et al. [3] implemented a system that understands itself to be speaking, by means of an efference copy of its intended speech action that can be compared to the incoming microphone signal.

We have highlighted some problems concerning spatial-self and self-reasoning in both language and vision. With few exceptions they have long been overlooked and still require more investigation. Representation and reasoning about self serves an underappreciated role in language, vision, and other cognitive abilities.

References

1. Miller, M.J.: A View of One's Past and Other Aspects of Reasoned Change in Belief. PhD thesis, University of Maryland at College Park, College Park, MD, USA (1993)
2. Brody, J., Cox, M.T., Perlis, D.: The processual self as cognitive unifier. In: Proceedings of the Annual Meeting of the International Association for Computing and Philosophy. (2013)
3. Brody, J., Perlis, D., Shamwell, J.: Who's Talking?–Efference Copy and a Robot's Sense of Agency. In: 2015 AAAI Fall Symposium Series. (2015)