Do Enactive Cognitive Robots Need a Self? Lessons from Neuroscience

Elena Antonova¹ and Chrystopher L. Nehaniv²
¹ Department of Psychology, Institute of Psychiatry, King's College London, UK
² Adaptive Systems Research Group, University of Hertfordshire, UK

Efforts in Artificial Intelligence, in particular and most promisingly, AI Enactive Cognitive Robotics seek to produce life-like agents by emulating the capacities seen in nature, particularly in humans. Human experience appears to involve several distinct notions of 'self', including somatic self, narrative self, social self, and experiential self. These 'selves', arising in ontogeny, are best thought of as processes that serve diverse and useful functions in humans. Their dynamic interplay and flexibility in shifting from one to another in a context-dependent manner is essential for adaptive (healthy) behaviour and cognition. The loss of such flexibility in humans can lead to various psychopathologies, such as depression and schizophrenia. Common to most human psychopathologies is the predominance of the narrative self-process, which on the psychological level is experienced as a self-entity separate from the other/environment. The narrative self has a distinct functional neural correlate in the human brain, the Default Mode Network (DMN), which runs in the midline of the brain, and minimally includes medial prefrontal cortex and the posterior cingulate. The ability to modulate and attenuate the DMN in humans is associated with greater behavioural flexibility, cognitive efficiency, emotional regulation, and general well-being.

Increasingly, AI robotics designers are finding they need to include mechanisms for narrative and social intelligence, introducing self-models, or targeting ‘self-awareness’. In the light of this, it seems that careful assessment is needed in determining what types of self-processes an enactive cognitive robot should embody and whether some modes of operating self-processes might reduce rather than increase functional and ‘metabolic’ efficiency.

In conclusion, in humans and hence, for similar reasons, in future enactive cognitive robots, ‘self’ is best conceived of as an interplay of transitory processes arising in a context-dependent and a function-specific manner. ‘Self’ as a continuously running process/module/network ‘overseeing’ the job of other processes in a top-down manner could be functionally inefficient and even detrimental.