Internal State Predictability as an Evolutionary Precursor of Self-Awareness and Agency

SfN 2008 (738.14)

November 19, 2008

Jaerock Kwon and **Yoonsuck Choe**

Department of Computer Science

Texas A&M University

For the full paper, see Kwon and Choe (2008).

Abstract

What is the evolutionary value of self-awareness and agency in intelligent agents? One way to make this problem tractable is to think about the necessary conditions that lay the foundation for the emergence of agency, and assess their evolutionary origin. We postulate that one such requirement is the predictability of the internal state trajectory. A distinct property of one's own actions compared to someone else's is that one's own is highly predictable, and this gives the sense of "authorship". In order to investigate if internal state predictability has any evolutionary value, we evolved sensorimotor control agents driven by a recurrent neural network in a 2D pole-balancing task. The hidden layer activity of the network was viewed as the internal state of an agent, and the predictability of its trajectory was measured. We took agents exhibiting equal levels of performance during evolutionary trials, and grouped them into those with high or low internal state predictability (ISP). The high-ISP group showed better performance than the low-ISP group in novel tasks with substantially harder initial conditions. These results indicate that regularity or predictability of neural activity in internal dynamics of agents can have a positive impact on fitness, and, in turn, can help us better understand the evolutionary role of self-awareness and agency.

Research Question: Self-Awareness

Why did self-awareness (or the sense of self) evolve?

- Self-awareness is an internal state that may be transparent to the process of evolution (cf. high-performance zombie).
- This is a hard question to answer without getting tangled in philosophical debate.

Strategy: Investigate the **necessary condition** of self-awareness that may be less controversial.