Investigating Objective Necessary Conditions of Consciousness in Simulated Evolution

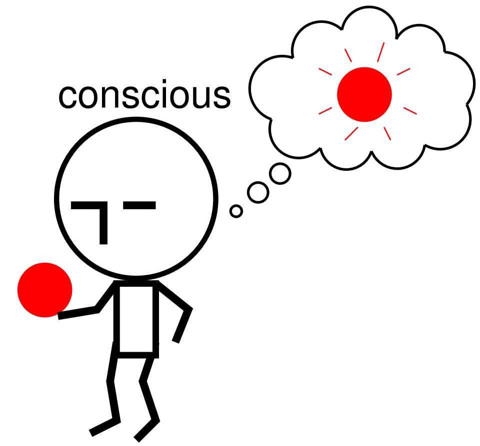
633 Machine Learning, Spring 2017

Yoonsuck Choe and Jaerock Kwon

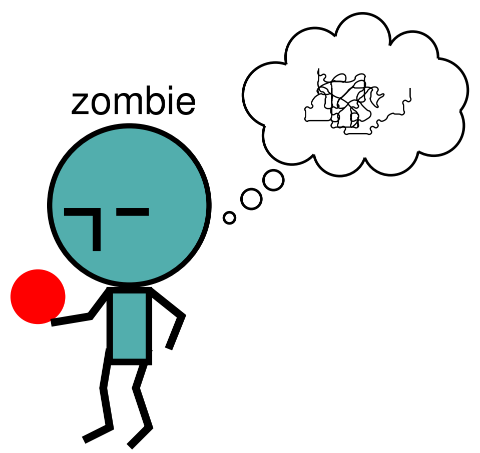
Department of Computer Science and Engineering

Texas A&M University

(Kwon and Choe 2008; Chung et al. 2009, 2012; Choe et al. 2012)



Function + Phenomenology



Function only – No phenomenology

I. Background and Motivation

What is Consciousness?

“Man, if you gonna ask, you’re never gonna know!”

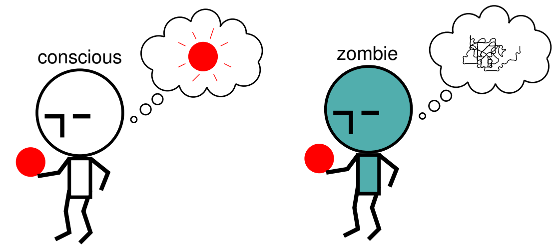
– Louis Armstrong, on jazz

- Definition is unclear, but even if definition is given, the problem can still be intractable.
- Both objective (functional) and subjective aspects (the “hard problem”).
- Objective and subjective aspects may be intertwined: This talk.

Approach and Immediate Problem

- Start with facts:

  Consciousness is a product of evolution.
- Immediate problem:
  - Evolution is blind to subjective phenomena, and it is hard to measure fitness in terms of such.

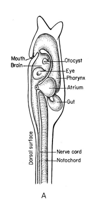


Potential Solution

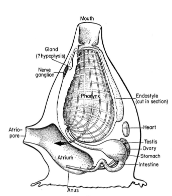
Potential solution: Investigate **objective necessary conditions** of consciousness.

- This can address the objective part (hard to measure degree of consciousness).
- But, what about the subjective part? – we will see how this unfolds!

Tree (no brain)



Tunicate (brain)



Tunicate (no brain)

cf. Llinás et al. (1994)

100% Predictable Future Event?

- Are there any future events that are 100% predictable?

100% Predictable Future Event?

- Are there any future events that are 100% predictable?
- What if I say there is such an event?

100% Predictable Future Event?

- Are there any future events that are 100% predictable?
- What if I say there is such an event?
- I'll wave my hand in the next 5 seconds.

100% Predictable Future Event?

- Are there any future events that are 100% predictable?
- What if I say there is such an event?
- I'll wave my hand in the next 5 seconds.
- "My" own actions are 100% predictable, but I can't predict your actios.

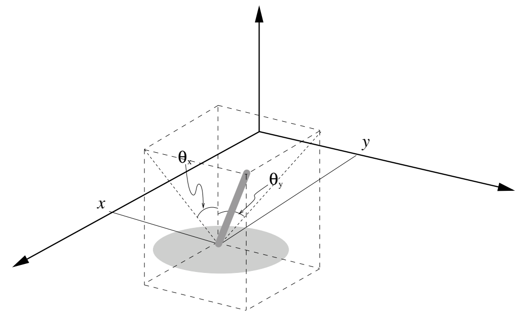
Main Insight

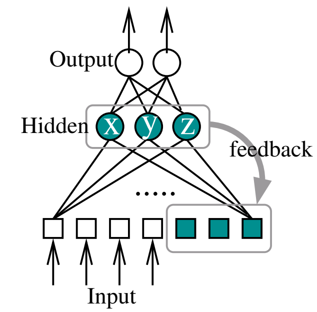
A meaningful necessary condition of consciousness is predictive capability:

1. Consciousness has a subject, the self.
2. Self is the author of his/her actions.
3. Property of self-authored actions (authorship): They are **100% predictable**!
   - Consciousness, self, and authorship are subjective, while predictability is objective.
   - Investigate predictive neural dynamics in simulated neuroevolution!

II. Methods

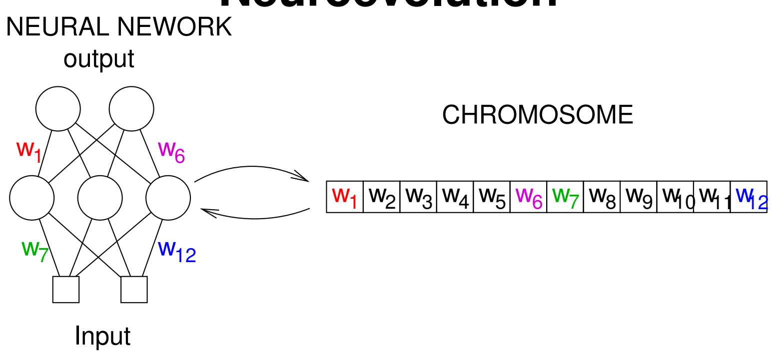
Overview





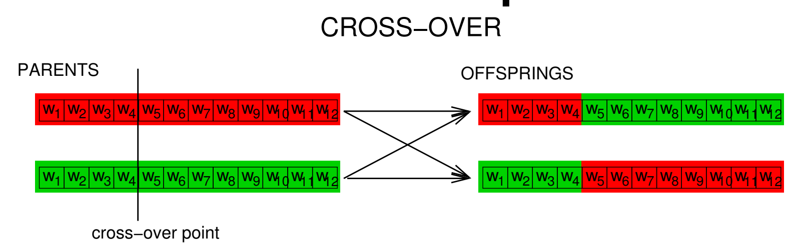
- Standard 2D pole balancing problem: Keep pole upright, within square bounding region.
- Evolve neural network controllers: Neuroevolution.

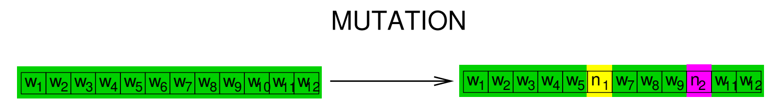
Neuroevolution



- A single chromosome encodes a full neural network.
- Each base pair, a single bit (or a real number), maps to a connection weight in the neural network.

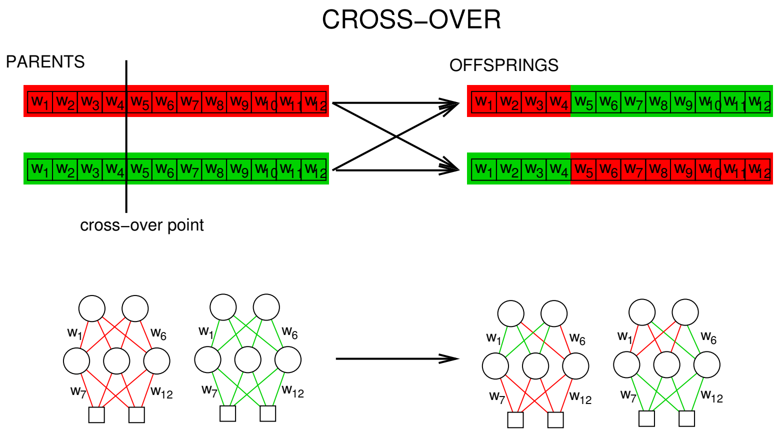
Neuroevolution: Operations





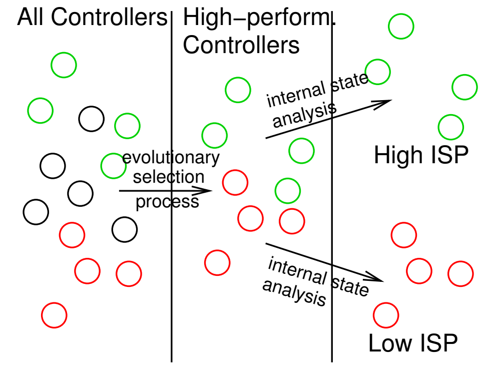
- Cross-over between two parents.
- Mutation of random base pairs.

Neuroevolution: Cross-Over, in Detail



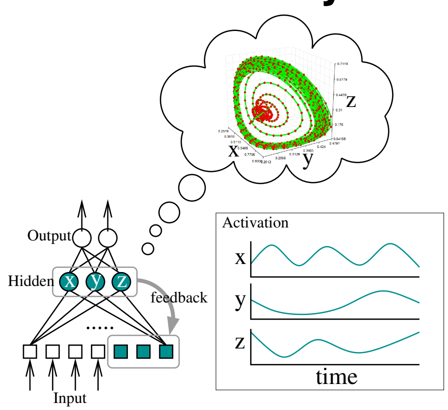
- Cross-over of two individuals produces two offsprings with a mixed heritage.

Overview of the Method



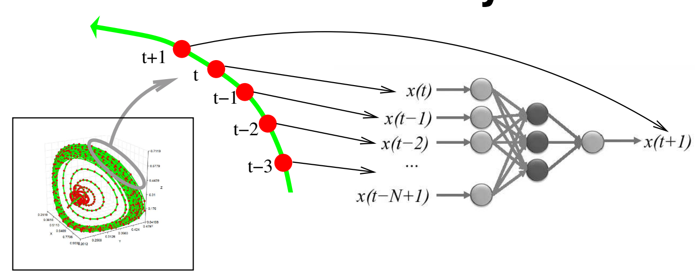
1. Evolve controllers to meet a fixed performance criterion (fitness does not measure predictability) in pole-balancing tasks.
2. Group high-performance individuals into high- and low internal state predictability (ISP) groups.
3. Test the two groups in harder tasks.

Internal Neural Dynamics



- Hidden layer activity can be thought of as the **internal state** of the neural network.
- Internal state dynamics: **Not directly observable**.

Measuring Internal State Predictability



- Train a **separate** feedforward network to predict the internal state trajectories.
- Measure prediction error made by the network.
  $\rightarrow$ High vs. low internal state predictability (ISP)

Method: Experimental Setup

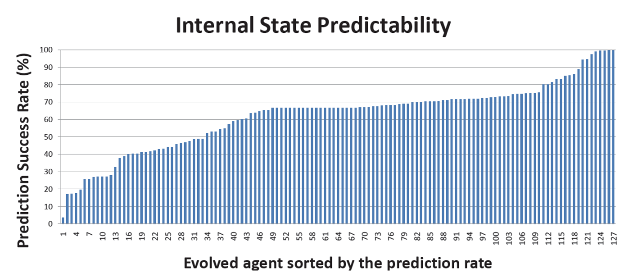
- Neuroevolution:
  - population size 50
  - mutation rate 0.2; cross over rate 0.7.
- 2D pole balancing task:
  - Pole should be balanced within 15° within a 3 m × 3 m arena.
  - Force applied to cart every 0.1 second (= one step).
  - Success if pole balanced over 5,000 steps.

Method: Experimental Setup

- Neural network predictor:
  - 2,000 training data.
  - 1,000 test data.
  - Back-propagation (learning rate 0.2).

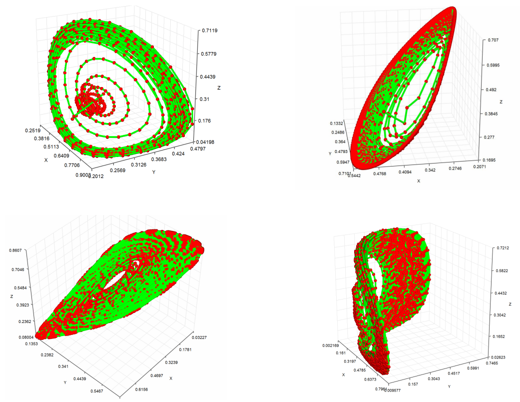
III. Results

Internal State Predictability (ISP)



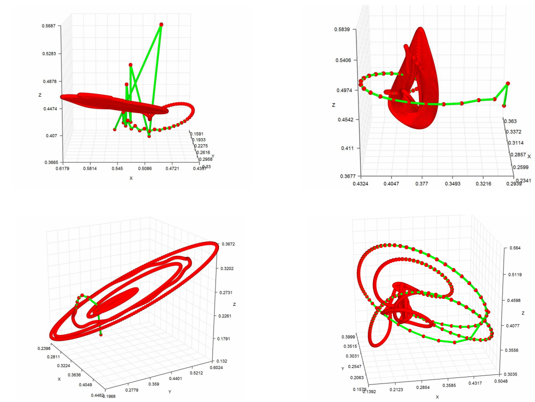
- Trained 130 pole balancing agents.
- Chose top 10 highest ISP agents and bottom 10 lowest ISP.
  - high ISPs: $\mu = 95.61\%$ and $\sigma = 5.55\%$.
  - low ISPs: $\mu = 31.74\%$ and $\sigma = 10.79\%$.

Examples of internal state dynamics from the high ISP group



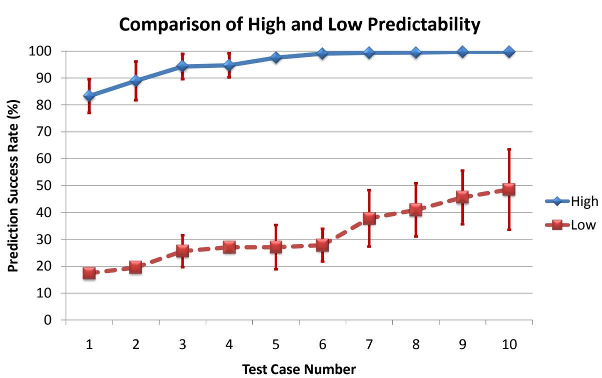
- Internal state dynamics show smooth trajectories.

Examples of internal state dynamics from the low ISP group



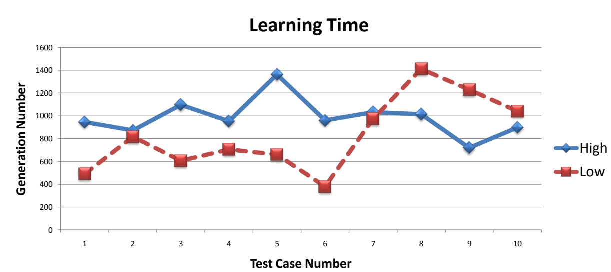
- Internal state dynamics show abrupt and jittery trajectories.

Comparison of High vs. Low ISP



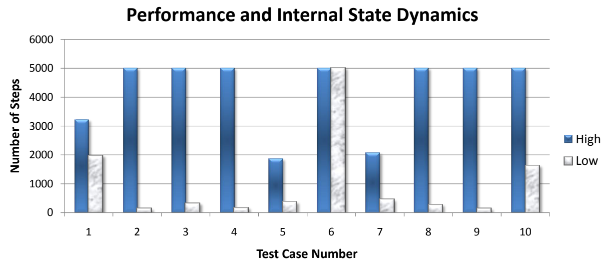
- A comparison of the average predictability from two groups: high ISP and low ISP.
- The predictive success rate of the top 10 and the bottom 10 agents.

Results: Learning Time



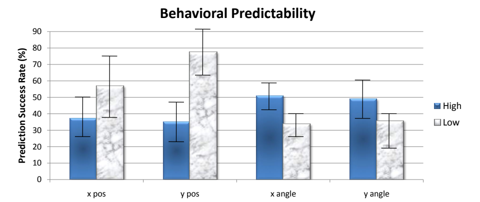
- No significant difference in learning time

Performance and Int. State Dyn.



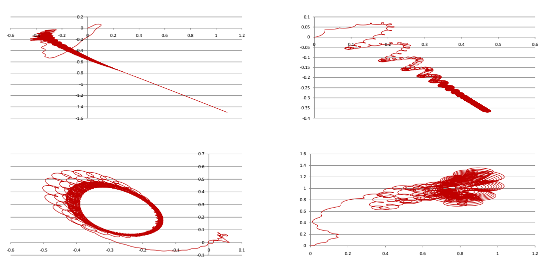
- Made the initial conditions in the 2D pole balancing task harder (pole more tilted).
- Performance of high- and low-ISP groups compared.
- High-ISP group outperforms the low-ISP group in the changed environment.

Behavioral Predictability



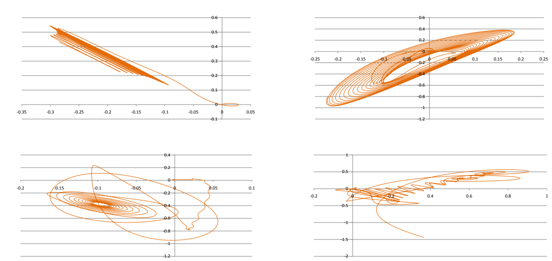
- Success of high-ISP group may simply be due to simpler behavioral trajectory.
- However, predictability in behavioral predictability is no different between high- and low-ISP groups.

Examples of cart x and y position from high ISP



- Behavioral trajectories of x and y positions show complex trajectories.

Examples of cart x and y position from low ISP



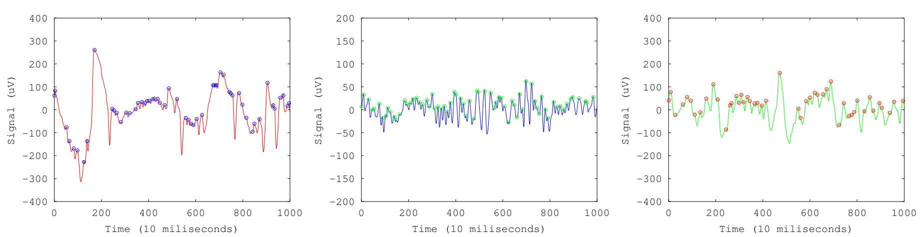
- Behavioral trajectories of x and y positions show complex trajectories.

IV. Brain EEG Analysis

Appilcation to the Real Brain?

- That's all just simulation: Can it be used in the real world?
- Yes!
  - Compare deep sleep vs. dreaming state vs. awake EEG.

Analysis of Real EEG Data

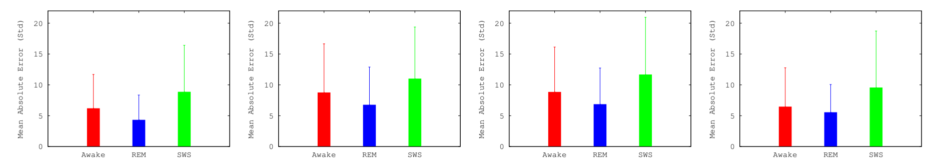


Awake REM Sleep Slow Wave Sleep

- Awake, REM sleep, and Slow-wave sleep EEG data.
- Inter-Peak Interval (IPI) predictability.

Yoo et al. *Frontiers in Neurorobotics* 2013.

Real EEG Data: Prediction Error



Subject 1 Subject 2 Subject 3 Subject 4

- All differences were significant ($p < 10^{-6}$) except for subject 4, Awake vs. REM.
- Bottom line: Awake and REM more predictable than SWS.

Yoo et al. *Frontiers in Neurorobotics* 2013.

V. Wrap Up

Discussion

- **Main contribution**: Found that an objective necessary condition of consciousness can evolve, despite the neural dynamics being internal.
- From subjective, to objective, back to subjective.
- Relavance of this work to the study of self-awareness, and the concept of self.
- Emergence of time in the brain: From the present (reactive, no memory), to the past (recurrent, with memory), to the future (predictive dynamics).

Discussion

- Imporatance of action in understanding consciousness (see e.g., Humphrey 1992; Llinás et al. 1994): Action and consciousness has many properties that go in parallel!
- Relationship between consciousness and grounding (of representations): See the famous Chinese Room Argument (Searle 1980) and action-based grounding (Choe and Smith 2006; Choe et al. 2007).
- Consiousness as a spatiotemporal phenomenon, and not as a "state" (cf. neural correlates of consciousness [Crick and Koch 2003])

Conclusion

- Internal state predictability has been identified as a necessary condition for consciousness.
- Predictive dynamics can evolve, in an increasingly competetive environment.
- **Lesson**: We can study subjective mental phenomena in a scientific manner by investigating non-trivial necessary conditions in the context of brain evolution.

Sources

- Tree: `http://homepages.inf.ed.ac.uk/jbednar/`
- Tunicate: `http://bill.srnr.arizona.edu/classes/182/Lecture-9.htm`

References

Choe, Y., Kwon, J., and Chung, J. R. (2012). Time, consciousness, and mind uploading. *International Journal on Machine Consciousness*, 4:257–274.

Choe, Y., and Smith, N. H. (2006). Motion-based autonomous grounding: Inferring external world properties from internal sensory states alone. In Gil, Y., and Mooney, R., editors, *Proceedings of the 21st National Conference on Artificial Intelligence(AAAI 2006)*, 936–941.

Choe, Y., Yang, H.-F., and Eng, D. C.-Y. (2007). Autonomous learning of the semantics of internal sensory states based on motor exploration. *International Journal of Humanoid Robotics*, 4:211–243.

Chung, J. R., Kwon, J., and Choe, Y. (2009). Evolution of recollection and prediction in neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, 571–577. Piscataway, NJ: IEEE Press.

Chung, J. R., Kwon, J., Mann, T. A., and Choe, Y. (2012). Evolution of time in neural networks: From the present to the past, and forward to the future. In Rao, A. R., and Cecchi, G. A., editors, *The Relevance of the Time Domain to Neural Network Models, Springer Series in Cognitive and Neural Systems 3*, 99–116. New York: Springer.

Crick, F., and Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6:119–126.

Humphrey, N. (1992). *A History of the Mind*. New York: HarperCollins.

Kwon, J., and Choe, Y. (2008). Internal state predictability as an evolutionary precursor of self-awareness and agency. In *Proceedings of the Seventh International Conference on Development and Learning*, 109–114. IEEE.

Llinás, R., Ribary, E., Joliot, M., and Wang, G. (1994). Content and context in temporal thalamocortical binding. In Buzsaki, G., editor, *Temporal Coding in the Brain*. Berlin: Springer Verlag.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–424.