References

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

International Journal of Machine Consciousness
© World Scientific Publishing Company

Time, Consciousness, and Mind Uploading

Yoonsuck Choe
Department of Computer Science and Engineering, Texas A&M University, 3112 TAMU, College Station, Texas, 77843-3112, USA. choe@tamu.edu

Jaerock Kwon
Department of Electrical and Computer Engineering, Kettering University, 1700 University Avenue, Flint, Michigan, 48504-6314, USA. jkwon@kettering.edu

Ji Ryang Chung
Department of Computer Science and Engineering, Texas A&M University, 3112 TAMU, College Station, Texas, 77843-3112, USA. jiryang@gmail.com. (Now at Samsung Electronics.)

Received: October 27, 2011
Revised:

The function of the brain is intricately woven into the fabric of time. Functions such as (1) storing and accessing *past* memories, (2) dealing with immediate sensorimotor needs in the *present*, and (3) projecting into the *future* for goal-directed behavior are good examples of how key brain processes are integrated into time. Moreover, it can even seem that the brain *generates* time (in the psychological sense, not in the physical sense) since, without the brain, a living organism cannot have the notion of past nor future. When combined with an evolutionary perspective, this seemingly straight-forward idea that the brain enables the conceptualization of past and future can lead to deeper insights into the principles of brain function, including that of consciousness. In this paper, we systematically investigate, through simulated evolution of artificial neural networks, conditions for the emergence of past and future in simple neural architectures, and discuss the implications of our findings for consciousness and mind uploading.

Keywords: mind uploading; time; material interaction; prediction; self; authorship; neuroevolution

1. Introduction

The function of the brain is intricately woven into the fabric of time. Functions such as (1) storing and accessing *past* memories [Shastri, 2002], (2) dealing with immediate sensorimotor needs in the *present* [Rossetti, 2003], and (3) projecting into, predicting, and/or anticipating the *future* for goal-directed behavior [Gross *et al.*, 1999; Henn, 1987; Kozma & Freeman, 2003] are good examples of how key brain processes are integrated into time. Moreover, it can even seem that the brain *generates* time (in the psychological sense, not in the physical sense) since, without the brain, a living organism cannot have the notion of past nor future (see Dowden

[2001] for a discussion on time and mind/brain function). When combined with an evolutionary perspective, this seemingly straight-forward idea that the brain enables the conceptualization of past and future can lead to deeper insights into the principles of brain function (see, e.g., Chung *et al.* [2009, 2012]; Chung & Choe [2009]; Kwon & Choe [2008]).

Most current investigations on the temporal aspects of brain function are focused on specific tasks such as temporal coding, binding/segmentation, or prediction (see, e.g., von der Malsburg & Buhmann [1992]; Fuhrmann *et al.* [2002]; Fortune & Rose [2001]; Natschläger *et al.* [2001]). Therefore, broader questions from an evolutionary perspective are rarely asked, e.g., can memory evolve from simple feedforward neural architectures, or can predictive function evolve from simple recurrent neural architectures? See Suddendorf & Corballis [2007] for a rare exception to this, where the authors talk about evolution of foresight and "mental time travel", albeit at a higher, cognitive level than what we will focus on in this paper.

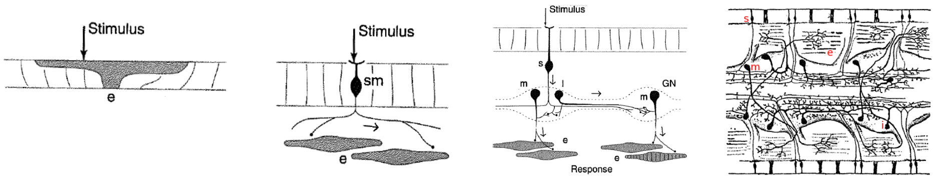
In this paper, we will systematically investigate, through simulated evolution of neural networks, conditions for the emergence of functions that enable the notion of past and future (i.e., memory and prediction) in simple neural architectures, and discuss the implications of the results for consciousness and mind uploading.

2. Background

In this section we will review works that provide background and motivation for this paper. First we will review the neural architectures of simple, primitive animals. Next, we will look at literature related to stigmergy (i.e., alterations of the environment to affect future behavior [Beckers *et al.*, 1994]) and its relation to memory. Finally, we will discuss literature on predictive functions in the brain.

2.1. *Neural architectures of primitive animals*

It is instructive to look at the origins of the nervous system and what steps it took to become the complex networks that we see today in advanced animals like mammals. Here, we will focus on the first few steps, shown in Fig. 1. One of the simplest animals is the sponge. The sponge lacks a nervous system, where independent effectors are actuated by direct stimulus (Fig. 1*a*). The next step up is the simplest animals with a nervous system, such as corals, jellyfish, and hydra. In these animals, e.g., in the hydra, a single neuron (sensorimotor neuron) links between the sensory surface and the effector (Fig. 1*b*) and these neurons form a sparse, distributed network. Finally, a more advanced form can be found in animals like the flatworm, where interneurons are introduced and cell bodies are organized into nervous ganglia along the length of the body (Fig. 1*c*-*d*). In all cases, the neuronal network of the animals are distinctly *feedforward*, thus their behaviors are largely reactive. Such animals can only respond to the moment-to-moment stimuli, oblivious of the inputs they received in the past (i.e., they live in the eternal present). In this sense, they do not have memory. Note that synaptic plasticity can be seen as a form of memory, but for simple animals like the flatworm, adjustment of synaptic efficacy can only change how the animal reacts to the immediate stimulus. How can such primitive animals further evolve to become sensitive to the past and the future? This is one of the central questions we will address in this paper.



(*a*) Single e (*b*) sm → e (*c*) s → m → e (*d*) Flatworm nerve net

Fig. 1. **Progression of neural architectures in simple animals** A progression of neural architectures is shown, from (*a*) a single effector (= e) that also senses the stimulus, to (*b*) a sensorimotor neuron (= sm) innervating the effectors (e.g., muscle cells), to (*c*) a sensory neuron (= s) innervating a motor neuron (= m), which in turn actuates the effectors. (*d*) Part of the flatworm nervous system showing a basic plan similar to (*c*) (i = interneuron, GN = ganglion). In all cases, the basic architecture is *feedforward.* Adapted from Swanson [2003]. (*d*) is from Cajal [1909], as shown in Swanson [2003].

2.2. *Stigmergy and its transition to memory-like function*

Stigmergy refers to "the production of a certain behaviour in agents as a consequence of the effects produced in the local environment by previous behaviour" [Beckers *et al.*, 1994]. Altering the environment in any way that affects future behavior falls under this category, such as dropping and detecting some type of marker. For example, humans and many other animals use external objects or certain substances excreted into the environment as a means for spatial memory (see Rocha [1996]; Chandrasekharan & Stewart [2004]; Chandrasekaran & Stewart [2007] for theoretical insights on the benefits of the use of inert matter for cognition). In this case, olfaction (or other forms of chemical sense) serves an important role as the "detector". Olfaction is one of the oldest sensory modalities, shared by most living organisms [Hildebrand, 1995; Vanderhaeghen *et al.*, 1997; Mackie, 2003]. This form of spatial memory resides in the environment, thus it can be seen as external memory. On the other hand, in higher animals, spatial memory is also internalized, for example in the hippocampus. Interestingly there are several different clues that suggest an intimate relationship between the olfactory system and the hippocampus. They are located nearby in the brain, and genetically they seem to be closely related: [Machold *et al.*, 2003; Palma *et al.*, 2004] showed that the Sonic Hedgehog gene controls the development of both the hippocampus and the olfactory bulb. Furthermore, neurogenesis is most often observed in the hippocampus and in the olfactory bulb, alluding to a shared functional demand [Frisén *et al.*, 1998]. Finally, it is interesting to think of neuromodulators [Krichmar, 2008] as a form of internal marker dropping.

Note that most existing works on stigmergy focus on the *social* aspect of it, such as in social insects and in ant colony optimization [Dorigo & Gambardella, 1997; Dorigo & Blum, 2005; Theraulaz & Bonabeau, 1999; Bonabeau *et al.*, 2000a; Carroll & Janzen, 1973], and in many cases they involve structure-building in the environment [Theraulaz & Bonabeau, 1999; Bonabeau *et al.*, 2000b]. In contrast, in this paper, we will present a more *individual* use of stigmergy, as a form of memory.

2.3. *Predictive function in the brain*

Prediction plays an important role in intelligent systems, and internal models make up a key component. For example, Rosen [1985] argued that anticipatory systems depend on internal predictive models of the agents themselves and the environments that are used in predicting the future for the purpose of control in the present. Wolpert and his colleagues showed how internal models in the brain (cerebellum, to be specific) can be used for prediction in motor behavior [Wolpert & Flanagan, 2001; Wolpert *et al.*, 1995, 1998; Kawato, 1999]. On the other hand, Bongard *et al.* [2006] showed that through the use of an internal self model, physical sensorimotor agents can show resilient behavior when part of the agent becomes damaged (such as amputated limbs, etc.). There are more instances of prediction being detected in brain function. Rao & Ballard [1999] showed that the interaction between feedforward and feedback connections between cortical regions can play a predictive role. Rao & Sejnowski [2000] also showed that predictive sequence learning can occur in recurrent cortical circuits. Finally, Hawkins & Blakeslee [2004] argued that the neocortex may have prediction as its primary function. In general, any work that cites anticipation, internal model, and goal-directed behavior all implicitly or explicitly involve prediction as a key part of their investigation.

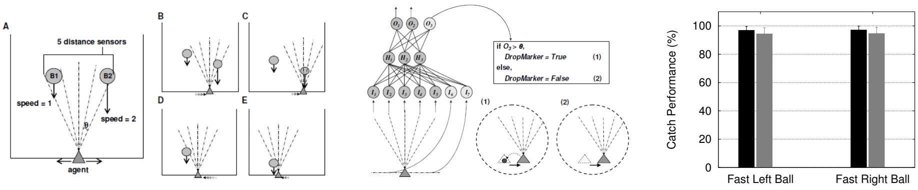
As mentioned in Sec. 1, most of the existing works on prediction focus on specific tasks or mechanisms, and rarely question the evolutionary origin of such a function. Furthermore, how memory (past) is related to prediction (future) is not considered in these works.

3. Emergence of Memory: From the Present to the Past

Can feedforward neural networks express memory-like behavior? In principle, this is not possible, but we found that when material interaction with the environment is allowed (basically a form of stigmergy), memory can be possible.

Fig. 2 summarizes the task, methods, and results. Fig. 2*a* illustrates the ball catching task [Beer, 2000]. Equipped with a fixed number of range sensors (radiating lines), an agent is allowed to move left or right at the bottom of the screen while trying to catch two balls falling from the top. The goal is to catch both balls. The balls fall at different speeds, so a good strategy is to catch the fast-falling ball first (Fig. 2*a* B and C) and then go back and catch the slow one (D and E). Note that in C the ball on the left is outside of the sensor range. Thus, a memory-less agent would stop at this point and fail to catch the second ball. In sum, this task requires memory.

Fig. 2*b* shows a feedforward network with a slight modification (dropper and detector of external markers). This kind of modification can be trivial to implement from an evolutionary point of view, since existing sensors can be extended to serve as detectors and excretion and other bodily discharge (e.g., pheromones) can take up the function of the dropper. The basic internal architecture of the network is identical to any other feedforward network, with five range sensor (I_1 to I_5), and two output units that determine the movement (O_1 and O_2). The two added input units (I_6 and I_7) signal the presence of a dropped marker on the bottom plane, and the additional output unit (O_3) makes the decision of whether to drop a marker at the current location or not. Note that there are no recurrent connections in the controller network itself. We used genetic search to evolve the network weights. The fitness was calculated based on the number of balls caught. The success in this kind of agent will depend critically on whether the markers are dropped at the right moment and appropriate behavior generated when certain markers are detected.



(*a*) Task (*b*) Dropper/detector network (*c*) Performance

Fig. 2. **Dropper/detector network's performance in ball catching** (*a*) The ball catching task is illustrated (A). The agent has five range sensors to detect the falling balls, and can move to the left or to the right. One ball falls fast and the other falls slow. To catch both balls, first the fast falling ball needs to be caught (B), but doing so would often make the slower ball to go beyond the sensor range (C). Memory is needed to go back to the slower ball (D–E). (*b*) The dropper/detector network is shown (it is basically feedforward). Five inputs indicate the range sensors for the falling balls, and two additional sensors are for detecting the markers on the ground. Two output units are used to determine the motion direction, and the third unit used to trigger the dropping of the markers. Genetic search is conducted on the connection weights of the controller network. (*c*) Performance comparison between the dropper/detector network (gray) and a recurrent network (black) is shown. See text for details. Adapted from Chung & Choe [2009, 2011].

In Fig. 2*c*, the average ball catching performance of the dropper network is presented (gray bar), along with that of the recurrent network (black bar). The recurrent network was a standard Elman network [Elman, 1991]. Both types of networks were trained using genetic algorithms, where the connection weights were adjusted over the generations (error bars indicate standard deviation). The results are reported in two separate categories: fast left ball and fast right ball. This was to show that the network does not have any fixed bias for catching the ball falling fast on one side only. Both networks performed at the same high level (over 90%

of the balls caught). This is quite remarkable for a feedforward network, although it had the added dropper/detector mechanism. The dropping/detecting strategy also seems consistent with an interpretation that the agent has memory (see Chung *et al.* [2009]; Chung & Choe [2011] for details). We also tested purely feedforward networks, but they were only able to catch only ∼50% of all balls dropped. The effectiveness of the dropper/detector network was also demonstrated in a more complex 2D foraging task [Chung & Choe, 2011].

These results suggest how organisms with only the concept of present could have evolved mechanisms to look into the past without rewiring their brain circuits.

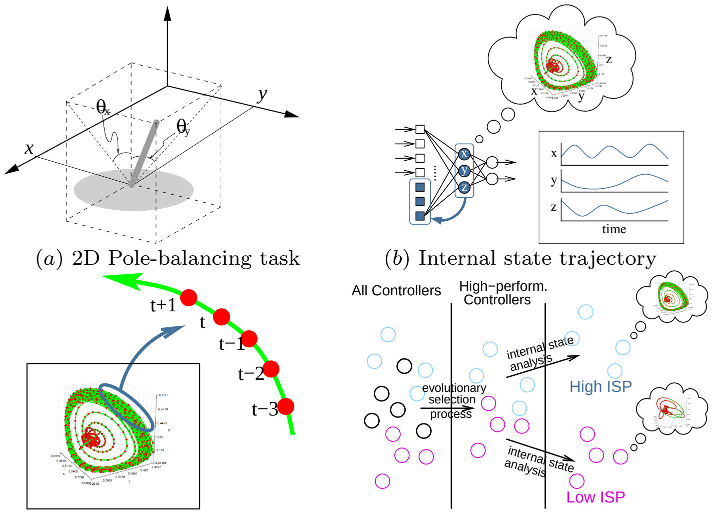
4. Emergence of Prediction: From the Past to the Future

Once a recurrent neural architecture is made available (through some route in evolution), what can it achieve? It can clearly modify its behavior based on stimuli that were received in the past. So, in a sense, recurrent neural networks have memory of the past, but that is only half of the story.

Do these recurrent networks have the ability to forecast the future? In fact, recurrent networks have been used extensively for time series prediction [Connor *et al.*, 1994; Barbounis *et al.*, 2006; Kuan & Liu, 1995]. However, these works are based on training the recurrent networks on time-series data that explicitly contain future information. Thus, the predictive capability emerging in these networks is mainly due to the information provided to them to begin with, through the supervised training set. That is, for this case, we cannot say that the predictive function was emergent.

We take a different stab at this question, by assuming no prior data that already contain information of the future, nor a built-in optimization criterion that explicitly measures prediction performance. The idea is to evolve recurrent neural network controllers in a dynamic task where prediction is not an immediate task requirement (Fig. 3*a*-*b*). The key innovation here was to simply measure the inherent predictability of the internal state trajectories (i.e., the time-series made up of hidden neuron activations over time; see Fig. 3*b*-*c*), and see if those with higher predictability spontaneously evolve, just based on the task demand (in this case, pole balancing). Note that here the predictability of the trajectory is a *post hoc* quantity only used for analysis, so it does not feed back into fitness calculation.

An immediate question is, how can predictability of the internal state trajectory be measured? Fig. 3*c* shows how. Again, we do not want to impose a preconceived notion (such as smoothness, low curvature, etc.) to bias our analysis, so we took a data-driven approach. Given n past data points ($n = 4$ in the figure), we want to know if the $n + 1$-th data point can be easily predicted. To measure this, we can construct a supervised learning training set consisting of n inputs and 1 target value, by sliding this small window along the entire internal state trajectory. Then, we can use any suitable supervised learning algorithm to learn the mapping, from n past data points to the $n+1$-th data point. Trajectories that lead to lowest training error can be said to have high predictability (i.e., the data set itself has a predictive property).



(*a*) 2D Pole-balancing task

(*b*) Internal state trajectory

(*c*) Measuring predictability

(*d*) Experimental design

Fig. 3. **Predictability of internal state trajectory in a pole-balancing controller network** (*a*) 2D pole balancing task. (x, y) is the location of the cart, and θ_x and θ_y the angles from the z axis. (*b*) A recurrent neural network controller for (*a*), illustration of its hidden-unit activations (internal state) over time (lower right, three neurons x, y, and z), and a 3D plot of the internal state trajectory. The connection weights are adapted using genetic search. (*c*) Measuring predictability of internal state trajectory. Given a few past data points as input ($t - 3$ to t), how well can the next data point ($t + 1$) on the trajectory be predicted? (*d*) Experimental design showing the population (left), selection (middle), and post-selection analysis (right). Individuals that pass the selection stage have equal task performance, but analysis of their internal state can show different characteristics: Some with highly predictable internal state trajectory, and others with much less predictable trajectory. (ISP = Internal State Predictability) Adapted from Kwon & Choe [2010].

Why should this be of interest? We have found that among equally high-performing individuals (Fig. 3*d*, middle), some have highly predictable internal state (i.e., hidden unit activation value) trajectories (Fig. 3*d*, right, top group and Fig. 4 top row [high ISP]) while some are not so predictable (low predictability, Fig. 3*d*, right, bottom group and Fig. 4 bottom row [low ISP]). In fact, the internal state predictability for 127 top-performing agents from the population show a smooth gradient, from very low to very high predictability (Fig. 5*a*). Since the individuals from both groups (high ISP vs. low ISP) passed the same performance threshold, they actually have equal performance given the same task. However, we discovered that when the initial condition of the task is made harder, those with high predictability retain their performance while those with low predictability lose much of their performance! (Fig. 5*b*)

The implication of this finding is profound. First, *predictable* internal state dynamics turned out to have a high selective value in evolution. This can be an important precursor to a full-blown predictive function. Second, certain properties that are *internal* to the agent can affect the course of evolution (examples of such internal properties include subjective phenomena such as consciousness: see Sec. 5 for more discussion on this point).

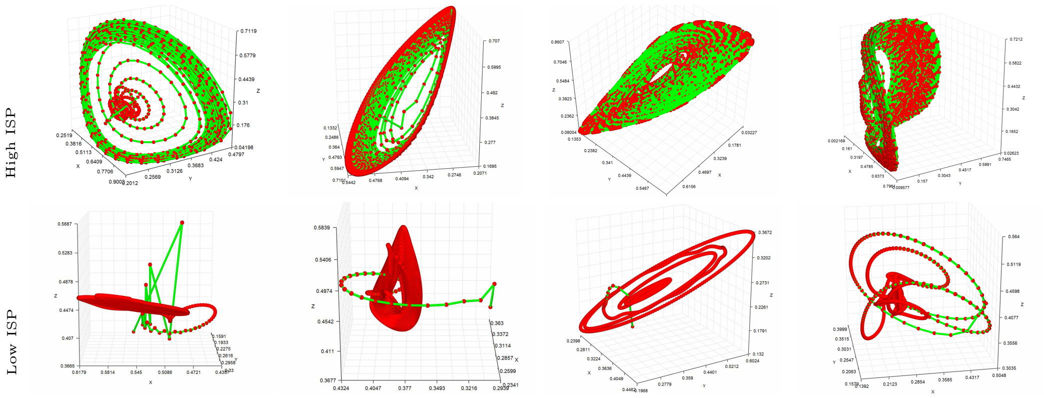
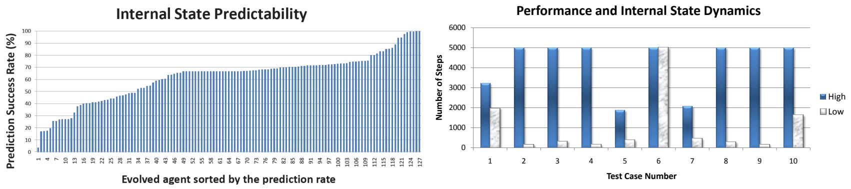


Fig. 4. **Internal State (Hidden Unit Activity) Trajectories** Examples are shown for highly predictable (top row) and hard to predict (bottom row) internal state trajectories. The highly predictable group shows smooth and periodic orbits, whereas the hard to predict group shows sudden turns and tangled trajectories. Adapted from Kwon & Choe [2008].

5. From Predictive Dynamics to Consciousness

Through our method outlined above, we can approach, in a scientific manner, one of the deepest mysteries in modern science, i.e., that of *consciousness* [Searle, 1997]. Here, we will talk about two aspects of consciousness that relate to the re-



(*a*) Internal state predictability (*b*) Performance in harder tasks

Fig. 5. **Internal State Predictability and Task Performance** Results from the analysis of internal state predictability and subsequent performance in harder task environments are shown. (*a*) The internal state predictability measured with a supervised learner is shown for 127 highly successful pole balancers. All of the controllers were able to balance the pole for 5,000 or more steps. (*b*) Comparison of the top 10 (blue bar) and bottom 10 (white bar) controllers in (*a*) are shown. In this comparison, the pole balancing task was made harder by increasing the initial tilt angle of the pole. We can see that the controllers with high internal state predictability mostly retain their performance, those with low predictability lose most of their performance. Adapted from results reported in Kwon & Choe [2008].

sults presented above (especially material from Sec. 4): (1) subjectivity and (2) self-perspectival organization [Van Gulick, 2004]. Subjectivity basically means that consciousness is a first-person property and is inaccessible by a third-person [Van Gulick, 2004]. This immediately raises questions regarding the evolutionary value of consciousness. For example, consider Fig. 6. If two equally functional individuals exist, one with subjective consciousness and the other without, why would natural selection favor the conscious one? Our results in Sec. 4 shed some light on this question. As shown in Fig. 3*d*, at some point in time, individuals with equal performance but with different internal properties (e.g., internal state predictability) can coexist. However, certain internal properties can at a later time (e.g., when the environment changes) turn out to be beneficial to survival. Thus, our work shows how *apparently* subjective properties can bias natural selection.

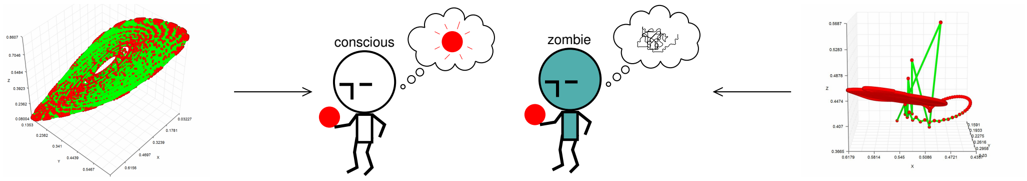


Fig. 6. **Conscious being vs. zombie** From the outside, a conscious being and a zombie (a philosophical zombie) may seem indistinguishable. However, internally (or subjectively), one might have phenomenal experience (left) while the other might lack this kind of experience (right). These internal characteristics may be determined in part by the internal state dynamics (see the text for details).

Self-perspectival organization means that conscious experience does not stand alone, but rather, it belongs to a subject or a self (Searle [1997], p. 183; Van Gulick [2004]). How can we scientifically study something that is so subjective as the notion of self? We propose that we need to track back by considering the properties and necessary conditions of these phenomena. For example, authorship (of one's own action) is a prominent property of the self. Authorship means "I" am the owner of my actions [Millgram, 2005; Humphrey, 1992]. A distinct property of self-authored actions is that they are 100% *predictable*. It does not make sense to say, e.g., "I think there is a 90% chance that I will type the letter 'A' after this." A necessary condition for such an accurate prediction is the predictability of the internal state dynamics, i.e., the underlying dynamics should lend itself to prediction (see Fig. 6). This is where it suddenly becomes objective, detached from the subjective notion of self. There is no agreement on how to measure the subjective aspect of consciousness or self, but we can experimentally (and hence objectively) measure the predictability of neural (or population) dynamics as a surrogate of the more subjective qualities of self and consciousness. Furthermore, our results in Sec. 4 showed that there is an evolutionary edge for agents that have more predictable internal dynamics. So, our work can provide clues on how important necessary conditions of consciousness could have evolved.

6. Implications on Mind Uploading

The basic idea of mind uploading is simple: Capture and preserve the mind state in a digital (or some other available) medium, for subsequent whole brain emulation [Hayworth, 2010; Sandberg & Bostrom, 2008; Koene, 2006]. However, there are diverse perspectives on what constitutes successful mind uploading and whether it is possible or not based on various theoretical and technical arguments [Hayworth, 2010]. In this section, we will discuss how the results presented in preceding sections can clarify issues regarding mind uploading relating to time, subjectivity, and evolution.

First, the temporal dimension needs attention equal to the spatial dimension when it comes to mind uploading (see the criticism of spatialization of time in Bergson [1988] which led to the concept of duration [durée], and the idea of "thick moment" by Humphrey [1992]). Structural (basically spatial) information alone, such as the connectome [Sporns *et al.*, 2005; Sporns, 2011], may not be sufficient for successful mind uploading and whole brain emulation. Of course parameters other than connectivity, such as connection strength and sign (excitatory or inhibitory), are important. However, the delay between connections are often overlooked, while our work suggests that delay can directly influence function since it can fundamentally alter the dynamics of the circuit, including the predictive kind. Furthermore, the existence of delay in the nervous system also seems to necessitate a predictive function, initially in the form of delay compensation [Lim & Choe, 2008, 2006a,b]. (See Choe [2004] for another example where temporal parameters [especially delay] play a key role in system-level function.) Thus, accurately estimating temporal parameters such as conduction delay based on the length and diameter of axons from structural data might be necessary for successful mind uploading and whole brain emulation.

Second, our approach suggests an effective strategy for dealing with subjective phenomena such as consciousness. As discussed in Sec. 5, instead of investigating the subjective phenomenon itself, we can initially focus on objective necessary conditions of the phenomenon. This way, important properties of the target phenomenon can be revealed and retained. In our case, we identified predictable dynamics as such a necessary condition. Such conditions (not just the ones we identified) can serve as a practical metric to measure the success of mind uploading and whole brain emulation. For example, two simulations that show the same performance on a task-specific metric can greatly differ in terms of their internal dynamics. In such a case, we can compare the internal dynamics to further refine our evaluation, e.g., preferring simulations exhibiting a more predictable internal dynamics.

Finally, our work demonstrates the importance of inferring the evolutionary steps through which mind emerged (cf. Humphrey [1992]). Mind uploading is in a sense very synchronic (as opposed to being diachronic), thus it may seem odd to put it in the context of an evolutionary time scale. However, as we have shown in this paper, considering an evolutionary perspective can help identify key principles that drive the emergence of mind. Such principles can effectively address theoretical issues regarding mind uploading such as subjectivity or self.

7. Discussion

In this paper, we investigated the relationship between time and consciousness, and identified predictable internal state dynamics as an important precursor of consciousness. Furthermore, we discussed the implications of this finding for mind uploading. In this section, we will examine three related topics: uncertainty, external vs. internal, and embodiment.

In Sec. 5, we argued that prediction of one's own action is an important precursor of consciousness. There is another angle from which we can link up prediction (especially those with high-confidence) and consciousness. The phenomenon of blindsight [Weiskrantz, 1986] is a good example. In blindsight, the subject is not consciously aware of the visual input, but visual task performance is maintained at a fairly high level. In a sense, the blindsight subjects are very uncertain, which could be indicative of low predictability leading to the lack of conscious perception. On the other hand, conscious subjects are extremely confident about their perceptual events. These observations allows us to see the three-way relationship among prediction, uncertainty, and consciousness.

Another interesting aspect of our work is that the boundary between internal and external (and similarly between subjective and objective) is blurred. Memory is needed, to eventually develop consciousness, but our first demonstration of memory was through the external medium (dropper/detector). The dropper/detector agent's neural architecture is feedforward, so it is supposed to be reactive, thus the agent lives in an eternal present. However, through the markers dropped in the environment, it gains memory. Thus, for this agent, the memory is straddled between the inner and the outer realm, blurring the subjective/objective boundary. Another example of such a blurring can be observed in the internal state predictability experiment. Can we say that the hidden unit activities are purely internal information inaccessible from the outside? That is, can we say that they are truly subjective? Experiments under harder task conditions was able to bring out agents that possess certain internal properties, so it seems hard to say the hidden unit information is totally subjective. These results challenge our preconceived notion of internal vs. external (or subjective vs. objective), and provide some conceptual tools for dealing with subjectivity in mind uploading research.

The last topic we want to discuss is embodiment. The focus of modern neuroscience is mostly on the brain, while the body receives relatively less attention [Hacker, 1987]. This trend is naturally quite dominant in mind uploading research, where the efforts are concentrated on preserving brain states. However, for successful whole brain emulation, we need to reverse engineer the brain, and for this, it will be necessary to map out the entire body as well, since it will be difficult to reconstruct the inputs and outputs without the body. Furthermore, important dynamics can only be captured at the level of the sensorimotor loop that involves the body and the environment (e.g., such as internal state invariance for autonomous semantics [Choe *et al.*, 2007]). In sum, for ultimate mind uploading, not only the brain but also the body may have to be scanned and emulated.

8. Conclusion

In this paper, we showed, through simulated evolution experiments, how memory of the past and prediction of the future can emerge in simple neural architectures. The results we find are simple yet profound: (1) even reactive mechanisms can gain memory, through actively altering the environment; (2) predictable internal state dynamics have a fitness advantage and will emerge through evolution when faced with changing environmental demands. Furthermore, we argued that (3) predictable dynamics is a precursor of consciousness. We expect these results to help us bring subjective issues in mind uploading into an objective domain, and provide concrete metrics and strategies for mind uploading, especially relating to the temporal dimension.

Acknowledgments

The main technical results summarized in this paper (Sec. 3 and 4) have appeared previously in [Kwon & Choe, 2008; Chung *et al.*, 2009; Chung & Choe, 2009, 2011; Chung *et al.*, 2012]. We would like to thank Timothy A. Mann (Texas A&M University) for his feedback on an earlier version of this manuscript.

References

Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C. and Dokopoulos, P. S. [2006] Long-term wind speed and power forecasting using local recurrent neural network models, *IEEE Transactions on Energy Conversion* **21**, 273–284.

Beckers, R., Holland, O. E. and Deneubourg, J. L. [1994] "From local actions to global tasks: Stigmergy and collective robotics," in R. Brooks & P. Maes (eds.), *Artificial Life IV*, Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems (The MIT Press).

Beer, R. D. [2000] Dynamical approaches to cognitive science, *Trends in Cognitive Sciences* **4**, 91–99.

Bergson, H. [1988] *Matter and Memory* (Zone Books, New York, NY), translated by Nancy Margaret Paul and W. Scott Palmer (from *Matiére et mémoire*, 1908).

Bonabeau, E., Dorigo, M. and Theraulaz, G. [2000a] Inspiration for optimization from social insect behaviour, *Nature* **406**, 39–42.

Bonabeau, E., Guérin, S., Snyers, D., Kuntz, P. and Theraulaz, G. [2000b] Three-dimensional architectures grown by simple 'stigmergic' agents, *Biosystems* **56**, 13–32.

Bongard, J., Zykov, V. and Lipson, H. [2006] Resilient machines through continuous self-modeling, *Science* **314**, 1118–1121.

Cajal, S. R. [1909] *Histologie due système nerveux de l'homme et des vertébrés, vol. 1* (Maloine, Paris).

Carroll, C. R. and Janzen, D. H. [1973] Ecology of foraging by ants, *Annual Review of Ecology and Systematics* **4**, 231–257.

Chandrasekaran, S. and Stewart, T. C. [2007] The origin of epistemic structures and proto-representations, *Adaptive Behavior* **15**, 329–353.

Chandrasekharan, S. and Stewart, T. [2004] "Reactive agents learn to add epistemic structures to the world," in K. D. Forbus, D. Gentner & T. Regier (eds.), *CogSci2004* (Lawrence Erlbaum, Hillsdale, NJ).

Choe, Y. [2004] The role of temporal parameters in a thalamocortical model of analogy, *IEEE Transactions on Neural Networks* **15**, 1071–1082.

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. and Choe, Y. [2009] "Emergence of memory-like behavior in reactive agents using external markers," in *Proceedings of the 21st International Conference on Tools with Artificial Intelligence, 2009. ICTAI '09*, pp. 404–408.

Chung, J. R. and Choe, Y. [2011] Emergence of memory in reactive agents equipped with environmental markers, *IEEE Transactions on Autonomous Mental Development* **3**, 257–271.

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* (IEEE Press, Piscataway, NJ), pp. 571–577.

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 A. R. Rao & G. A. Cecchi (eds.), *The Relevance of the Time Domain to Neural Network Models, Springer Series in Cognitive and Neural Systems 3* (Springer, New York), pp. 99–116.

Connor, J. T., Martin, R. D. and Atlas, L. E. [1994] Recurrent neural networks and robust time series prediction, *IEEE Transactions on Neural Networks* **5**, 240–254.

Dorigo, M. and Blum, C. [2005] Ant colony optimization theory: A survey, *Theoretical Computer Science* **344**(2-3), 243–278.

Dorigo, M. and Gambardella, L. M. [1997] Ant colony system: A cooperative learning approach to the traveling salesman problem, *IEEE Transactions on Evolutionary Computation* **1**(1), 53–66.

Dowden, B. [2001] Time, in J. Fieser & B. Dowden (eds.), *Internet Encyclopedia of Philosophy*, `http://www.iep.utm.edu`. Retrieved December 13, 2010.

Elman, J. L. [1991] Distributed representations, simple recurrent networks, and grammatical structure, *Machine Learning* **7**, 195–225.

Fortune, E. S. and Rose, G. J. [2001] Short-term synaptic plasticity as a temporal filter, *Trends in Neurosciences* **24**, 381–385.

Frisén, J., Johansson, C. B., Lothian, C. and Lendahl, U. [1998] Central nervous system stem cells in the embryo and adult, *CMLS, Cellular and Molecular Life Science* **54**, 935–945.

Fuhrmann, G., Segev, I., Markram, H. and Tsodyks, M. [2002] Coding of temporal information by activity-dependent synapses, *Journal of Neurophysiology* **87**, 140–148.

Gross, H.-M., Heinze, A., Seiler, T. and Stephan, V. [1999] Generative character of perception: A neural architecture for sensorimotor anticipation, *Neural Networks* **12**, 1101–1129.

Hacker, P. [1987] Languages, minds and brain, in C. Blakemore & S. Greenfield (eds.), *Mindwaves: Thoughts on Intelligence, Identity, and Consciousness*, chap. 31 (Blackwell, Oxford, UK), pp. 485–505.

Hawkins, J. and Blakeslee, S. [2004] *On Intelligence*, 1st ed. (Henry Holt and Company, New York).

Hayworth, K. [2010] Killed by bad philosophy: Why brain preservation followed by mind uploading is a cure for death, Essay published online at `http://www.brainpreservation.org`.

Henn, V. [1987] Cybernetics, history of, in R. L. Gregory (ed.), *The Oxford Companion to the Mind* (Oxford University Press, Oxford), pp. 174–177.

Hildebrand, J. G. [1995] Analysis of chemical signals by nervous systems, *Proceedings of National Academy of Sciences, USA* **92**, 67–74.

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

Kawato, M. [1999] Internal models for motor control and trajectory planning, *Current Opinions on Neurobiology* **9**, 718–727.

Koene, R. A. [2006] Scope and resolution in neural prosthetics and special concerns for the emulation of a whole brain, *The Journal of Geoethical Nanotechnology* **1**, 21–29.

Kozma, R. and Freeman, W. J. [2003] Basic principles of the KIV model and its application to the navigation problem, *Journal of Integrative Neuroscience* **2**, 125–145.

Krichmar, J. L. [2008] The neuromodulatory system: A framework for survival and adaptive behavior in a challenging world, *Adaptive Behavior* **16**, 385–399.

Kuan, C.-M. and Liu, T. [1995] Forecasting exchange rates using feedforward and recurrent neural networks, *Journal of Applied Econometrics* **10**, 347–364.

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* (IEEE), pp. 109–114.

Kwon, J. and Choe, Y. [2010] "Predictive internal neural dynamics for delay compensation," in *Second World Congress on Nature and Biologically Inspired Computing (NaBIC2010)*, pp. 443–448.

Lim, H. and Choe, Y. [2006a] Compensating for neural transmission delay using extrapolatory neural activation in evolutionary neural networks, *Neural Informa-*

tion Processing–Letters and Reviews **10**, 147–161.

Lim, H. and Choe, Y. [2006b] "Facilitating neural dynamics for delay compensation and prediction in evolutionary neural networks," in M. Keijzer (ed.), *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, GECCO-2006*, pp. 167–174.

Lim, H. and Choe, Y. [2008] Extrapolative delay compensation through facilitating synapses and its relation to the flash-lag effect, *IEEE Transactions on Neural Networks* **19**, 1678–1688.

Machold, R., Hayashi, S., Rutlin, M., Muzumdar, M. D., Nery, S., Corbin, J. G., Gritli-Linde, A., Dellovade, T., Porter, J. A., Rubin, S. L., Dudek, H., McMahon, A. P. and Fishell, G. [2003] Sonic hedgehog is required for progenitor cell maintenance in telencephalic stem cell niches, *Neuron* **39**, 937–950.

Mackie, G. O. [2003] Central circuitry in the jellyfish aglantha digitale iv. pathways coordinating feeding behaviour, *The Journal of Experimental Biology* **206**, 2487–2505.

Millgram, E. [2005] Practical reason and the structure of actions, in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Stanford University, Stanford, CA), [online] `http://plato.stanford.edu/entries/practical-reason-action/`.

Natschläger, T., Maass, W. and Zador, A. [2001] Efficient temporal processing with biologically realistic dynamic synapses, *Network: Computation in Neural Systems* **12**, 75–87.

Palma, V., Lim, D. A., Dahmane, N., Sánchez, P., Brionne, T. C., Herzberg, C. D., Gitton, Y., Carleton, A., Álvarez Buylla, A. and Altaba, A. R. [2004] Sonic hedgehog controls stem cell behavior in the postnatal and adult brain, *Development* **132**, 335–344.

Rao, R. P. N. and Ballard, D. H. [1999] Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects, *Nature Neuroscience* **2**, 79–87.

Rao, R. P. N. and Sejnowski, T. J. [2000] "Predictive sequence learning in recurrent neocortical circuits," in S. A. Solla, R. K. Leen & K.-R. Muller (eds.), *Advances in Neural Information Processing Systems 13* (MIT Press, Cambridge, MA), pp. 164–170.

Rocha, L. M. [1996] Eigenbehavior and symbols, *Systems Research* **13**, 371–384.

Rosen, R. [1985] *Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations* (Pergamon Press, New York).

Rossetti, Y. [2003] Abstraction from a sensori-motor perspective: Can we get a quick hold on simple perception? *Philosophical Transactions of the Royal Society of London Series B* **358**, 1269–1275.

Sandberg, A. and Bostrom, N. [2008] Whole brain emulation: A roadmap, Tech. Rep. #2008-3, Future of Humanity Institute, Faculty of Philosophy and James Martin 21th Century School, Oxford University.

Searle, J. [1997] *Mystery of Consciousness* (The New York Review of Books, New York).

Shastri, L. [2002] Episodic memory and cortico-hippocampal interactions, *Trends in Cognitive Sciences* **6**, 162–168.

Sporns, O. [2011] *Networks of the Brain* (MIT Press, Cambridge, MA).

Sporns, O., Tononi, G. and Kötter, R. [2005] The human connectome: A structural description of the human brain, *PLoS Computational Biology* **1**, e42.

Suddendorf, T. and Corballis, M. C. [2007] The evolution of foresight: What is mental time travel and is it unique to humans, *Behavioral and Brain Sciences* In press.

Swanson, L. W. [2003] *Brain Architecture: Understanding the Basic Plan* (Oxford University Press, Oxford).

Theraulaz, G. and Bonabeau, E. [1999] A brief history of stigmergy, *Artificial Life* **5**, 97–116.

Van Gulick, R. [2004] Consciousness, in E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Stanford University, Stanford, CA), [online] `http://plato.stanford.edu/entries/consciousness/`.

Vanderhaeghen, P., Schurmans, S., Vassart, G. and Parmentier, M. [1997] Specific repertoire of olfactory receptor genes in the male germ cells of several mammalian species, *Genomics* **39**, 239–246.

von der Malsburg, C. and Buhmann, J. [1992] Sensory segmentation with coupled neural oscillators, *Biological Cybernetics* **67**, 233–242.

Weiskrantz, L. [1986] *Blindsight*.

Wolpert, D. M. and Flanagan, J. R. [2001] Motor prediction, *Current Biology* **11**, R729–R732.

Wolpert, D. M., Ghahramani, Z. and Jordan, M. I. [1995] An internal model for sensorimotor integration, *Science* **269**, 1880–1882.

Wolpert, D. M., Miall, R. C. and Kawato, M. [1998] Internal models in the cerebellum, *Trends in Cognitive Sciences* **2**, 338–347.