Consciousness, edited by P. Zelazo, M. Moscovitch, and E. Thompson, 117–50. New York, NY: Cambridge University Press, 2007.

O'Regan, K. "Explaining What People Say about Sensory Qualia." In *Perception, Action, and Consciousness*, edited by N. Gangopadhay, M. Madary, and F. Spicer, 31–50. Oxford University Press, 2010.

Pereboom, D. *Consciousness and the Prospects of Physicalism*. Oxford University Press, 2011.

Schwitzgebel, E. (2014) "Introspection." *The Stanford Encyclopedia of Philosophy*, summer 2014 edition, edited by Edward N. Zalta. http://plato.stanford.edu/archives/sum2014/entries/introspection

Shoemaker, S. "The Inverted Spectrum." *Journal of Philosophy* 79 (1982): 357–81.

———. "Self-Knowledge and 'Inner Sense': Lecture I: The Object Perception Model." *Philosophy and Phenomenological Research* 54, no. 2 (1994): 249–69.

Sloman, A., and R. Chrisley. "Virtual Machines and Consciousness." *Journal of Consciousness Studies* 10 (2003): 113–72.

Smith, A. *The Problem of Perception*. Cambridge, MA: Harvard University Press, 2002.

Turing, A. "Computing Machinery and Intelligence." *Mind* 59 (1950): 433–60.

Tye, M. *Ten Problems About Consciousness*. Cambridge, MA: The MIT Press, 1995.

———. "Qualia." *The Stanford Encyclopedia of Philosophy*, Fall 2015 edition, edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2015/entries/qualia/

From Biological to Synthetic Neurorobotics Approaches to Understanding the Structure Essential to Consciousness, Part 1

Jeffrey White

KOREAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (KAIST) COMPUTATIONAL NEUROSYSTEM LABORATORY, DEPARTMENT OF ELECTRICAL ENGINEERING, DRWHITE@KAIST.AC.KR

Jun Tani

KOREAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (KAIST), DEPARTMENT OF ELECTRICAL ENGINEERING

ABSTRACT

Direct neurological and especially imaging-driven investigations into the structures essential to naturally occurring cognitive systems in their development and operation have motivated broadening interest in the potential for artificial consciousness modeled on these systems. This first paper in a series of three begins with a brief review of Boltuc's (2009) "brain-based" thesis on the prospect of artificial consciousness, focusing on his formulation of h-consciousness. We then explore some of the implications of brain research on the structure of consciousness, finding limitations in biological approaches to the study of consciousness. Looking past these limitations, we introduce research in artificial consciousness designed to test for the emergence of consciousness, a phenomenon beyond the purview of the study of existing biological systems.

SECTION 1: INTRODUCTION

Nature seems here eternally to impose a singular condition, that the more one gains in intelligence the more one loses in instinct. Does this bring gain or loss?

– Julian Offray de La Mettrie[1]

The following paper is the first of three. It sets out the case for research in artificial consciousness, arguing that studies in artificial systems are a necessary complement to research into biological systems due both to the nature of artificial systems as well as the limitations inherent in studies of biological systems. First, it briefly introduces Piotr Boltuc's "naturalistic non-reductionist" account of consciousness which holds that "first person consciousness is not reducible to material phenomena, but that it is at the same time fully explainable by such phenomena."[2] Then, the second and third sections of this paper explore some of the implications of studies into biological consciousness, one of which being that the "pure" subjectivity that is the object of some philosophical discourse is quickly occluded by concomitant processes and overlapping networks. Through the discussion, Boltuc's originally clear assay gives rise to two more complex types of consciousness, **most-consciousness** and **myth-consciousness**, both apparently necessary and not accidental aspects of human cognitive agency. We find a complimentary account in recent work from Thomas Fuchs, and here are met with practical limits to consciousness research in biological systems. In the third section, we follow Edelman and Baars in looking directly at research into artificial consciousness as a way past these limitations. Finally, the fourth section quickly reviews a series of experiments establishing the emergence of a minimal self-consciousness in lead up to the second paper in this series, which reviews this group's most recent work on freewill.

Concerning artificial consciousness, Boltuc has issued a positive thesis. He is confident that artificial consciousness is possible when the material nature of biological cognition is better understood. "Machines can be conscious like any organism can."[3] He offers an analysis of consciousness into three forms, functional, phenomenal and h-consciousness ("hard"), and he raises questions about a locus of consciousness based on existing biological systems.

On Boltuc's estimation, robots are already what he calls "functionally" conscious. Through their normal function, "they can perform many thinking tasks comparable, or superior, to humans, though by other means."[4] "Thinking" for Boltuc is simple enough, being "any kind of information processing that increases inductive probability of arriving at a correct result"[5]—i.e., error correction. So, thinking is integral to learning. Phenomenal consciousness is more complex, and at the center of what Boltuc takes to be "*the* most important, but somewhat neglected, philosophical issue in machine consciousness today", that "every function attributed to p-consciousness could, in principle, be played by an AI mechanism using some sort of functional mechanism, only."[6] That this is not yet the case is due specifically to the lack of an adequate "generator of consciousness" the functions of which, once understood adequately, will be able to be engineered.[7]

Boltuc analyzes p-consciousness into subcategories, the "broad" also "functional" sense indicative of "first-person functional consciousness" including direct perception, and the "narrow" "non-functional" sense indicative of the "mineness" that characterizes human-like "hard" "h-consciousness". H-consciousness is the focus of Boltuc's engineering thesis[8] because it represents the "awareness" of being, "the locus of first-person experiences", and he argues that without this awareness "there is nothing that it is like to be that robot."[9] Important to Boltuc's analysis here is his distinction between subject and object. A subject is ultimately a non-object, and an object a non-subject. For Boltuc, this constitutes the simplest ontology, and helps to further clarify the special nature of h-consciousness. One way to understand the first-person perspective is as that "subjective perspective from which one performs a certain function (e.g., the perspective from which one makes a picture)" but another way is "the very stream of awareness that a conscious individual has."[10] One is "inside" and the other, the former, remains a third-person perspective on the first-person perspective. For Boltuc, this distinction underscores the difference between a proto-cognitive system like a camera, or a robot with some minimal degree of "consciousness", and the different case that is h-consciousness. He labels it the "Is anybody home?" problem essential to "systems with their own locus of awareness."[11]

The "is anybody home" problem has to do with feeling the difference between before and after states consequent on thinking actions, as an agent experiences and necessarily (perhaps permanently) embodies what at least seems to follow from conscious phenomena. Being able to answer the "Is anybody home?" question is the reason that h-consciousness is "at least a condition of one's status as a moral patient strictly understood."[12] And, given also "a strong, plausible tendency to view moral value as dependent on first person awareness (h-consciousness)", the ability to answer this question carries "strong implications for ethics and in particular for the relative moral standing of robots, as they are now, and animals (including humans)."[13] After all, Boltuc is not ready to afford moral status to entities simply because they are h-conscious, e.g., "rats."[14] More seems to be needed, and we will begin to look into what this more might amount to in the next section.

Finally, Boltuc argues that an artificial consciousness is unlikely to emerge as an aspect of a computer "program". On his assay, a program can model complex biological systems, but that "they are not those systems" and therefore "it is very unlikely that h-consciousness is merely a feature of a program."[15] His advice is to pursue inquiries into biological systems first, understand for example what differentiates human cognition from that of a rat, and from this end "try to build a generator of consciousness in some other, inorganic or organic, matter or, if possible, find them in some already existing systems" i.e., as "generated in human and other animal brains."[16] And, this is where we are left, with the challenge to both conceive of how h-consciousness can be fully explained in material terms through an understanding first of how h-consciousness is "generated" in available biological systems and then, with this understanding, to "engineer" it. This is Boltuc's non-reductive naturalistic thesis.

There are some immediate problems with such a proposal. For one thing, success is effectively impossible to confirm due to the fact that the "verification of results" is confounded given the privileged access that characterizes h-consciousness as the "mineness" of experience. The work now is to account for this mineness in the most direct way. Consider Michel Bitbol's view.[17] On Bitbol's estimation, subjectivity is not something extra, it is essential to cognition, for consciousness, yet so is objectivity and the result is a necessary "dance" between first and third person perspectives in the representation of consciousness as a "stabilized and intersubjectively shared structural residue."[18] Problems with privileged access to subjective states and the problem of other minds remain imperfectly resolved, but this is the nature of conscious systems and to be expected. Such is not the nature of artificial conscious systems, however, and from them we may form different expectations.

The hurdle of privileged access may be overcome with perfect information about the dynamic structure of a cognitive system ready at hand, a perspective not afforded human observers of natural cognitive systems *in situ* such as those which are Bitbol's and Boltuc's main concern. This potential is afforded, however, by artificial systems as we shall see in greater detail going forward. First, we must further establish the limits of the use of biological models in the search for a formal structure of consciousness.

SECTION 2: TEMPORALITY

At issue is the potential for conscious machines, specifically artificial systems with a sense of ownership over their actions and intended ends. Piotr Boltuc has advised that research on the hard problem of engineering artificial consciousness should focus on the structure of "existing systems" in order to understand consciousness as it may be made to exist in non-human artificial agents, specifically through engineering "projectors of first-person awareness."[19] Consider this fact in approaching the problem of consciousness as posed in Chalmers' zombie thought experiment.[20] A focus on structural isomorphism doesn't seem very promising in solving the hard problem of consciousness in zombies, as these are structurally identical with existing conscious systems. On the form of Chalmers' thought experiment, consciousness must be something other than structural isomorphism at the finest grain of material assay.

Chalmers' zombies help to spotlight the fact that with every reduction of consciousness into material nature there remains the question, what is missing in a zombie equivalent. This is the "hard" problem of consciousness. There are different ways of trying to zero out the debt that remains on the "full" explanation of consciousness in purely material terms.[21] One may confess to being a zombie. One may posit the existence of a locus of the feeling of mineness of consciousness, typically some *organelle* and corresponding operations within a biological brain without which consciousness in whatever form is impossible, which is the general direction recommended on Boltuc's thesis as well.[22] Some lines of inquiry isolate consciousness to networks of activity at the center of which is a hub of activity in the thalamus, with ongoing work in the structure internal to the thalamus and how this anatomy correlates with consciousness.[23] This is not a unique view; it is not unpopular and not new given the longstanding recognition of the thalamus as a special hub of neural activity central to consciousness.[24]

However, if we are to look at isolating a distinct region of central neural activity as the locus of h-consciousness in particular, then the thalamus may not be the best candidate area. After all, Boltuc's engineering thesis merely advises that any effort at artificial consciousness should aim to recapitulate something performing as a "projector" or "generator" of consciousness, and this role might be played by a number of candidate systems. One possibility is the reticular activating system, for example. The reticular formation sits at the confluence of the internal environment of the central neural system above it and the external perceptual reality as mediated by the body system below. And, its role in the "projection" of consciousness is well-known, for example as set out by Parvisi and Damasio who argue that consciousness arises when an organism is able to "internally construct and internally exhibit a specific kind of wordless knowledge" that "the organism has been changed by an object … along with the salient enhancement of the object image caused by attention being allocated to it."[25] Consciousness then arises as the agent adjusts to the experience, a process enabled by the reticular formation and carried forward by the reticular activating system.[26]

That said, many regions above the reticular formation seem to be even more important to the sense of "mineness" characterizing h-consciousness in particular. When searching for the locus of the feeling of what it is to be "me" rather than another subject, consider the ventromedial prefrontal cortex which—as Bechara, Damasio and Damasio describe[27]—"links" particular perceptions with established emotional valences. Moreover, these associations are then modulated, especially reinforced post-choice in the reduction of internal inconsistencies ("cognitive dissonance") and without awareness.[28] Folded into the discussion thus far, it is difficult to imagine what could be more "mine" than the surprising feeling of an unexpected adjustment to a disposition to act. And yet, the vmPFC is involved in processing specific to other essential ingredients of the mineness of experience, as well.

If anything were more mine than the felt update on prior embodied yet hidden preferences, then it may be one's anticipations of a personal future. Damage to the ventromedial region correlates with the inability to take up anticipatory emotional states as evidenced by skin conductance on the presentation of a decision situation, with subjects optimizing for short rather than long term rewards, "oblivious to the future."[29] Other research has demonstrated that reduction in activity of the vmPFC correlates with reduced predictive capacity due to the fact that the vmPFC enacts processes that effectively populate possible futures from the first-person perspective such that a failure in predictive capacity ultimately derives from the "failure to think self-referentially about our future selves."[30]

In short, due to activity in the vmPFC, we may say that a human cognitive agent "has a future", and moreover we may say that this future is essentially social with deep moral implications for a biologically realistic account of h-consciousness, as well. The vmPFC is implicated in "empathic decision making" which involves making decisions in order to optimize another's future well-being.[31] Accordingly, vmPFC damage has been associated with impaired moral emotions such as empathy central to morality and implicated in moral judgment.[32] The right vmPFC especially is implicated in empathy, with damage to this area resulting in, among other things, reduced moral sensitivity to situations involving perceived injustice.[33] The central thesis here is that evolved biological drives result in "moral emotions" that in the vmPFC automatically conjoin self and other interests in constraining possible futures towards which cognitive agency is then exercised. The result is the creation of joint attention and "intersubjective space" as the default form of future into which a self is projected in part through vmPFC processes (and echoing Bitbol in an interesting way). Taken as a whole, this research affords insight into the essentially social nature of human cognition due the essential social nature of the generation of the possible future self through activity in the vmPFC in particular.[34] Human cognitive agency is social agency, simply put.[35]

Here, we find a locus of activity contributing to the mineness of h-consciousness that is at the same time essentially social and also temporal, outstripping Boltuc's original analysis of h-consciousness as pure subjectivity. And as we explore the implications of this activity, the original analytic sense of zombie has also mutated into something more, something closer to actual human beings, perhaps moral zombies instead. After all, Boltuc intends merely that h-consciousness is a necessary but insufficient condition for being a moral patient, as a "locus of awareness" characteristic of first-person experience. Yet, the mineness characteristic of h-consciousness as we have been developing it, in consideration of neural processes and how these are bundled, reveals the essentially social and temporal dimensions of what we may call "**most-consciousness**" (mine other self temporal) instead of merely h-conscious in order to differentiate from Boltuc's analysis.

vmPFC processing makes it a prime candidate as a locus of most-consciousness for an essentially social cognitive agent, perhaps especially when integrated with the dmPFC into the entire mPFC.[36] After all, if anything were more characteristic of the "mineness" of experience than the surprising adjustment to erstwhile hidden preferences and redirection of one's future project self, it may be the empathy opened to others, directly allowing one's self and its projected future to be emotionally transformed through moral perspective taking. With this in mind, then, an easy answer to "What is the most-zombie missing?" is "a future" or perhaps "a future with friends in it" or perhaps just as well, a vmPFC. Finally, if we accept as essential the relationship between a project future central to mineness and moral agency exercised toward this internally constructed end then it stands to reason that one way to engineer a zombie without a sense of ownership of its own agency is to somehow interfere with the function of the thalamus, or with the vmPFC, to take its present, or its future, or both.[37]

But, what about its past? Time consciousness involves not only future and present, but also past. Nothing may be more "mine" than my own future, and how I feel about it, except perhaps my own past, how this brings me into the present and disposes[38] me toward some futures rather than others. Without a past, one may be sensitive to changes without recognizing the difference between before and after as if on a perpetual roller coaster with no time to think. Likewise, we may imagine that zombies may be without pasts, without memories, without the mineness that characterizes most-consciousness.

Memory formation is thought to depend mostly on another area of the brain, the hippocampus, and interfering with the function of the hippocampus can result in something like a zombie. One interesting and more or less common loss of most-consciousness corresponds with the loss of memory in an alcohol blackout. The mineness of consciousness is lost along with the feeling of before and after. During blackouts, affected individuals often execute complex action routines including speech and the use of symbols within noisy and even dangerous environments while being left with often spotty memories which seem to indicate that p-consciousness was in some limited way present, but lost.[39] Of the rest, there is no sense of mine-ness. There is no memory. Something here is missing, Boltuc's "very stream of awareness" is interrupted, and this is what makes an alcohol blackout like being a zombie.[40]

When we think most broadly about the constituents of a unique self, especially about what is unique to this agent as opposed to any other, we might be drawn to the notion of memory. The vmPFC is necessary, and the thalamus, certainly, and all are structurally and functionally unique to each subject at the finest grains of analysis, but without a memory of how one used to feel about something, before that preference changed, then the "mineness" characteristic of most-consciousness is also impossible.[41] In a way, then, the hippocampus seems to be a good location on which to focus if one were intent on the creation of most-zombies, i.e., beings exactly like us but without most-consciousness. However, it is not a difference in neural anatomy that makes the difference here. Rather, it is the presence of magnesium ions within an ion channel that modulates memory formation. Blackouts happen when Mg++ doesn't get into the channel to block the influx of ions because without this plug, impaired memorization and long-term brain damage result.[42] Thus, to the question "What is the zombie missing?" one may answer "Magnesium ions in receptor channels modulating NMDA receptor function" rather than name any neuroanatomical organelle.

Moreover, if we can imagine a drug which keeps these channels open to an influx of ions that results in the burnout of memory formation – perhaps permanently and reliably given certain selective stimuli – then we can imagine the purposeful creation of zombies which are potentially selectively incapable of consciousness of certain things and relations. The field of perception can be stabilized by unconscious processes, and so a stable subject is estimable, but in fact the feeling of mineness about one's own direct experiences would be absent without a sense of change at least seemingly dependent on conscious processes and moderated by past conditions which are absent, on this rather soft zombie example.[43] Without this before and after, we are left with a mindless doppleganger due an alcohol blackout, a being with most-conscious potential but without the memory that partly constitutes the sense of self about which most-consciousness is concerned.

Even this construct does not help us to solve the hard problem as originally formulated in Chalmers (1996) because at the finest grains of analysis it is not structurally identical with a non-zombie. Due to the presence or absence of ions and other neurotransmitters, molecular conformations change. Potentials for development change. Futures change, and even disappear. But, that doesn't mean that we can't get clearer on the relationship between consciousness and memory by holding the alcohol blackout alongside the zombie model. To be a zombie requires that the subject first have the potential for most-consciousness, and then to be denied its realization. The question is after all in the form of a "What is missing?" And, memory certainly qualifies as a natural non-reductionist candidate, fully explainable but not fully reducible to material description, after all being dependent on context and interpreted for others including one's self in reflective inner discourse. At the same time, there is a strong association between memory and moral agency as illustrated in the fact that human beings may be exempted from misdeeds performed during blackout states that would otherwise invite greater sanction.[44]

The flip-side of the problem of other minds is the issue of accounting for one's own. How much must be accounted for, and what is the best way to do it? How much memory does a cognitive agent need in order to be conscious in a morally relevant way, not be snuffed out as a nuisance? More than a rat? What kinds of memories are necessary? What kind of future is necessary? And, if these are all necessary, then isn't the hippocampus also necessary along with the ventromedial and the thalamus? Where with a biological model of consciousness must we stop for an adequate account of consciousness? With the brain of the agent? Its skin? The systems in terms of which it is embedded?

Here, especially, we can see the role for symbolic expression in the construction of narratives that make conscious exposition possible. Symbols help us to remember. And they also help us to project. From ink and paper to the printing press, the first popular fictions were psychological self-reports.[45] With these narratives as subjective, first-person anticipatory and regretful accounts of life from the inside-out so to speak, there is the modern sense that what is important is not determined and the past perhaps best left behind, with the future open and at least potentially within control, the modern project which so given represents simply a ubiquitous aspiration intersubjectively distilled.[46] We will have something to say, in the third paper, about the motivational potential arising as horizons of anticipation are projected due normal adolescent neural development.

In the end, if articulating artificial consciousness means simulating all of this complexity in a computational medium, e.g., artificial cognitive agents which write papers for publication on the prospects of artificial cognition, then we may well have before us an impossible task.

SECTION 3: MYTH-CONSCIOUSNESS

We have been sorting out how to understand biological models of consciousness in a way which affords a neat view on especially the "mineness" of consciousness which we have since developed from an analytic shell into a biologically more realistic sense of most-consciousness. But what is the relationship between consciousness and cognition generally speaking? In his exposition, Boltuc defines "cognition" as "interactions of a system with an environment" and importantly he adds a requirement, that this system must "involve structural retention of some pieces of information". On this "somewhat zoocentric" account, for both biological and robotic "organisms" cognition can be construed without "reference to consciousness, as processing of sensory input and its assimilation into existing schemes" as the agent "gains knowledge" and "becomes aware" and then "uses that knowledge for comprehension and problem solving."[47] Cognition is very much like "thinking" which on Boltuc's recipe works essentially to solve problems. The essential difference is that structural retention alone such as that involved in "learning" remains "short of consciousness" along with "sleepwalking" until "down towards simpler organisms", e.g., "roaches," questions for example about moral status due to consciousness become meaningless "since we seem to have no reason to presume consciousness apart from those [unconscious cognitive] processes."[48] Thus, Boltuc offers a definition of consciousness in relation to cognition – that it is a "special instance" of cognition, which cannot be reduced to simple cognition, yet which operates as an extension of embodied cognitive agency, the exercise of which ideally opens further opportunities for continued cognition, i.e., as the agent becomes aware of problems worth solving.

Given that problems facing agents may arise in all modes of said agency, any realistic account of the "mineness" of consciousness is unlikely to limit itself to any single region or sub-process within the brain. Is it possible, then, that we may account for consciousness in terms of more distributed neural systems? Consider Edelman, Gally and Baars on what consciousness requires:

> Consciousness consists of a stream of unified mental constructs that arise spontaneously from a material structure, the Dynamic Core in the brain. Consciousness is a concomitant of dynamic patterns of reentrant signaling within complex, widely dispersed, interconnected neural networks constituting a Global Workspace.[49]

"Unified" phenomena become so by the harmonic coordination of networks via the "Dynamic Core" on the thesis that consciousness arises due to cortical processes as they re-enter the thalamus from various regions. As we have seen, the thalamus is the nexus of "reentrant signaling" produced by the complex, widely dispersed, interconnected neural networks which constitute Baar's (1997) "Global Workspace" some of which is engaged within any given task environment.[50]

Dynamic core models clarify a number of things. For example, the subjective sense that consciousness is something over and above simple cognition is reinforced in the anatomy of thalamocortical loops as more connections develop between the thalamus and the frontal areas of the cortex in the direction of the thalamus than the other way around. And, on this more systematic view, we find that consciousness arises in the synchronization of distributed operations rather than in virtue of one class of cells within one sub-region of the central nervous system. That said, though such an account may tell us why a thalamus is necessary for consciousness in biological systems like ours, it does not tell us in what form it may be essential to conscious systems in general.[51]

Recognizing such limitations in biologically reductive approaches to consciousness, Edelman, Gally and Baars recommend that "A theory of human consciousness... must rest on a more global theory of how vertebrate brains are organized to yield function."[52] And this means exactly looking past the structure internal to the brain itself for the influences shaping central neural system organization.

In this light, consider Thomas Fuchs' effort at understanding cognitive agency as enabled deeply embodied material memorization, with the agent as a whole its own record, and this again only significant in light of an agent's projected and anticipated future selves.[53] This is a "more global" account than those reviewed so far, as it begins with autopoietical self-organization and identifies consciousness with processes set on maintaining the integrity of the organism in the face of disintegrative change. Fuchs sees cognitive agency as "integrative" embodiment, with memory distributed both within the brain and also in the material processes of the distributed body system as it internalizes the world in its interactions. Consciousness, thus, emerges in the "diachronic unity" of cognitive agency, as the agent sets out and feels its distance from more or less ideal situations with this proto-natural inclination ultimately shaping how brains and bodies are organized to yield function.

"The systemic unity of the organism thus becomes the precondition of the unity of self-experience"[54] as the "diachronic unity of consciousness" is formed by "a self-referential process in which each succeeding moment implies an awareness of the next-to-come and the just-past" which results in "a pre-reflective self-awareness"[55] of the imminence of these instances as they are to be embodied. In accounting for this process, Fuchs stresses the role of self-organization of an organism individuated not by accident but due to its embodied nature and by way of which there arises the experience of "continuity of the self from a first-person perspective" the "pre-reflective feeling of sameness or a felt constancy of subjectivity" to which one awakens prior to any remembering or reconstruction of an object self.[56] And, he finds the substrate of this felt sameness in the concept of "bodily existence" itself.

Bodily existence is characterized first of all by the "diachronic coherence of a basic bodily self" and this coherence specifies only "an abstract identity or sameness . . . but no qualitative identity," i.e., it is purely formal. It tells us nothing of "the sort of persons that we are" and it is for this same reason that it is unrealistic, neglecting the fact that "all enactments of life are integrated into the memory of the body, and here they remain preserved as experiences, dispositions, inclinations and skills."[57] In other words, on Fuchs' account, consciousness is always and already of deeply embodied material memory with this record also establishing implicit valuations on experience that are more or less malleable (e.g., in the case of an octopus, not so much). In so far as this embodied agency is furthered or hindered, healed or injured things are good or bad, and the body as memory is the record of this status quo as well as how to deal with it.[58] "Body memory is thus the ensemble of all habits and capacities at our disposal."[59]

Rather than looking for a generator of conscious phenomena, Fuchs finds the grounds of consciousness in the temporal structure of consciousness, the binding together not of subjective with objective points of view, but rather future with past and not within a self as a separable process, but constitutive of self (and likely demanding a "multi-layered" understanding of self[60]). Working from a Husserlian analytic, Fuchs writes that "the stream of experience as a continuous synthesis of what is not-yet, what is now, and what is no-longer" constitutes the "diachronic unity of consciousness" as a "self-referential process in which each succeeding moment implies an awareness of the next-to-come and the just-past" resulting in, again, the "pre-reflective self-awareness"[61] that is the also the target of Boltuc's h-consciousness. However, it is only when coupled with the deep material memory of embodied cognition that this "pure" subjectivity takes on its unique character, mine own such that "without its embedding in the continuity of pre-reflective bodily existence" the mineness of consciousness disappears and "the narrative self and its memories remain but [as] a story that we tell about an alien person."[62]

It is not simply a matter of occupying a position in a course of historical evolution that is at issue, here. Rather, the capacity for the subject to employ embodied resources in the direction of this history is a difference that is worthy of distinction from the simple model of most-consciousness that we have developed thus far. Consider here why Edelman, Baars and Seth hold forth for the necessity of narrative facilitated by language in the exposition of the "mineness" characteristic of human consciousness as it facilitates the detachment of the subject from the feeling of being its self in the present.[63] This same capacity allows human beings to represent for others similarly embodied the series of conformations undergone in a felt, embodied transformation from one situation to another, e.g., we can learn from others' self-reported experience, and reflect on our own in the same way. On the other hand, the authors do not find this capacity in octopi as they appear unable to adapt neural structures driving goal achievement in recognizable response to contextual cues in a laboratory environment, so seemingly making any narrative progress beyond simple evolutionary forces impossible and any question about how octopi might communicate changes to others moot (for example, through skin color changes). As this case illustrates, there is a distinction to be made between a cognitive agent acting within the space of its evolution, determined by its inherited form, and a cognitive agent with the capacity to *make* this history both through symbolically represented narrative exposition as well as through self-directed self-change, becoming the agent required for the execution of some action or other the necessity of which arising first in the subjective projection of possible future self-situations of and for that very agent. This new form we may call "**myth-consciousness**." This distinction between what we have been calling "most-" and "myth-" consciousness is worth some attention.[64] The latter's important role comes largely from the fact that cognitive agents more or less embody the vastness of history and its determinations.[65]

Myth-consciousness retains the mineness and essentially moral social temporal nature of most-consciousness, but recognizes that as embodiment is the medium of memory, and that as embodied memory is history, then the nature of (human-like) consciousness is essentially historical as well. From this point of view, "Pure consciousness without a subjective body is a dualistic abstraction which forgets that all thinking owes its emergence to the preceding process of life."[66] These life processes are the assemblage of continuously unfolding problem solving routines unique to a uniquely historically situated, uniquely materially embodied social cognitive agent, i.e., Boltuc's "thinking."

Fuchs writes that "through our habits, we inhabit the world"[67] with this habitation constraining focal attention, while also cementing the agent into the landscape of living history which is its focus. Thus we may stress that this is a circular process. The world informs (as in "puts the form into") our habits, and through these habits we change the world that again informs our habits upon which we then act.

Emphasizing the historical depth of embodied memory, Fuchs points to the difference between traumatic and more everyday memories with interesting implications for our understanding of myth-consciousness, as well. In illustration of the way that trauma affects embodied memory, Fuchs quotes Aahron Applefield who survived as a refugee from Ukraine under fascism during the last World War: "The cells of my body apparently remember more than my mind which is supposed to remember."[68] And, Fuchs is right to draw attention to the difference in deep memory of traumatic versus everyday events. Jovasevic et al. have shown in a mouse model two pathways for memory encoding active in the hippocampus, one which distributes memory to higher level cortical regions, and another which passes traumatic memories to sub-cortical systems essentially outside the reach of conscious introspection.[69] Surprisingly, the same GABA receptors in the hippocampus which had been associated with the impairment of fear-related memorization facilitate the retrieval of fear-related memory states, and they do this by promoting subcortical activity as opposed to distributed cortical activity typically associated with episodic memory. Of course this makes sense. It serves the survival of an agent to respond reflexively to dangers in the environment, with the somatic marker of fear or rage attaching to those biochemical changes resulting from pre-cortical information processing. This is the pre-reflexive condition into which we awake, and on Fuchs' account this level of embodied memory especially grounds personal identity. Thus, the more or less stable subject over time that, as a relatively regular pattern of activity that "emerges . . . from a history of embodied experience which has accumulated and sedimented in body memory and as such is implicitly effective in every present moment"[70] is more or less constituted by unconscious processes, and this has serious implications for any consideration of the "mineness" of consciousness.[71]

The ghost is not in the machine. The ghost *is* the machine. Troubles arise when embodied habits do not suit the changing environment, when embodied existence cements history in its "bones" and the traumatized agent can no longer adapt habits to a different habitat. At this extreme, there is no longer any ghost, only mechanism perhaps amounting to a kind of zombie. And here, with the pre-reflective capacity to adjust to environmental changes in the maintenance of prospective integrity, Fuchs points out that these "circular processes" of self regulation are "arguably necessary for the emergence of basic self-awareness" within an artificial consciousness as well. An artificial consciousness must find itself situated in the world, on its way to different situations, with prospective self-awareness and also memory about these situations and the transitions between them.[72]

With this, Fuchs brings us to a practical limit facing any program in the study of biological consciousness. Every stage of development of an organism embodying Fuchs' deep material memory—and experiencing those peculiarly "mine" moments, for example when a human being completes a paper on consciousness after years of conscious and unconscious preparation—cannot be reliably isolated in the study of a biological system. They may be reported, shared. But, they cannot be exactly reproduced. All of the dimensions weighed in the use of one term rather than another, for example, cannot be systematically tracked. This is not the case for inquiries into artificial conscious systems, however. Indeed, recognizing a similar limit to the biological approach, Edelman, Gally and Baars issue something of a compliment to Boltuc's engineering thesis, advising that we must "accept" that we cannot map cognition in the study of living beings, "to trace causal chains at all levels of complexity in the brain circuits that contribute to consciousness" while at the same time suggesting that a "brain-based device, driven by a simulated brain . . . would be key to success" in understanding consciousness, instead.[73]

In summary, where Edelman, Gally and Baars recommend research into "brain-based" devices, and Boltuc likewise points to "generators" of consciousness within biological brains, Fuchs suggests that the prospects for artificial consciousness emerge at the interface of embodied cognitive agent and environment, at the level of whole organism in the social historical temporal world that is mine and yours in so far as we embody these horizons. So, to the question "What is a zombie missing?" one might answer "Itself" as a whole.[74] In the next and final section of this paper we introduce a research program in neurorobotics which instantiates "circular processes" such as those which Fuchs finds requisite for basic self-awareness, leaving the next paper in this series to set out in detail this groups' work in freewill and self-reflective consciousness.

SECTION 4: MINIMAL SELVES

While Boltuc cuts the cognizer into two logical aspects, subject and object, Edelman, Gally and Baars emphasize a dynamic core within a global workspace, and Fuchs finds the subjective and objective standpoints to be together essential to cognition in integrative embodied agency. One thing that all share is a positive assessment of the prospects of a properly configured artificial consciousness, and all generally agree on how such a machine might be built, replicating part or all of an organic system. Following such a recipe in an artificial medium faces difficulties with replicating processing dynamics due to biochemical reality. We will approach these issues in the third paper in our series, when we revisit Boltuc's natural non-reductionism. Current technology does not afford computational power to simulate realistic human brain activity. However, we may not need to instantiate whole brains and narrative-historical political consciousness in chemical metabolisms with all attending systems due natural embodiment in order to isolate aspects of consciousness. Rather, specific features might be drawn in their essential dynamics, such that "a much smaller number of simulated neurons and synapses might prove sufficient to give rise to a particular mental property, such as imagery."[75]

One particularly important aspect of the problem of consciousness as we have drawn it in discussion thus far is the problem of time consciousness, or "temporality", and one especially interesting aspect of temporality is how the raw flow of perceptual experience is parsed and consolidated into narratives composed of sequences of events involving objects as well as other subject agents.[76] We will describe recent experiments involving the instantiation into robots of this capacity to construct and to deconstruct possible futures, to aim for them so as to explore the consequences in the next paper. Here, by way of introduction, we will briefly review how this research program demonstrates the emergence of "basic self-awareness" in the form of a minimally self-reflective self.

Tani's basic model employs higher and lower levels of differently configured neural networks with the latter tuned to the immediate environment and responsive to rapid changes while the former higher level is attuned to longer ranged patterns. It is in the interactions between these two levels that Tani finds consciousness arising, and he has spent the last two decades building robots which demonstrate this to be the case (for most complete review, see Tani, in press). Here, discussion should turn to the notion of predictive coding and its relationship with the diachronic unity of Tani's neurorobots.[77]

In 1998, Jun Tani detailed a dynamical system structure accounting for the phenomenon of the momentary appearance of the "self" and demonstrated these dynamics in robotics experiments. Tani showed that the self emerges momentarily when the coupled dynamics between the internal neural network and the environment shift from coherent to incoherent dynamics. When everything proceeds as anticipated in the coherent phase, there is no distinction between the self and the environment in the coupled dynamics. However, the self can be perceived as separate from the environment when something goes wrong, in conflict with the system's anticipation, in the incoherent phase.

In the first experiment (Tani, 1998), constructive and deconstructive interactions between the bottom-up pathway of perception and the top-down pathway of prediction were balanced by internal parameters derived from prior prediction error. Throughout the learning process, the entire system dynamics proceeded with intermittent shifting between the coherent and incoherent phases, with good predictability in the former and poor predictability in the latter. These results were interpreted though Heidegger's (1996) famous example of the hammer, i.e., we become aware of the use of the hammer only when the hammer fails to perform as anticipated, such as when it breaks. In this case, Tani postulated that the gap generated between top-down anticipation and bottom-up reality in the incoherent period represent the difference between the unconscious, routine use of a hammer and its perhaps violent mechanical failure. In this moment, Tani conjectured, the structure of cognitive agency as a "minimal self" rises to awareness. As the agent looks for "What went wrong?" it takes itself as a possible object and answer, "I went wrong." Further, Tani conjectured that the entire system dynamics tends to proceed toward a certain critical state in which a large range of fluctuations may take place, a condition analogous to a system at criticality.[78]

Tani and Nolfi (1999) and Tani (2003) further explored this problem of self-referential selves.[79] Especially, in a learning experiment with a robot navigating a maze environment (Tani & Nolfi, 1999) and one with a robotic arm manipulating an object (Tani, 2003), the continuous sensorimotor flow of information became segmented into reusable behavior primitives. This chunking was accomplished through a dynamic gate opening/closing (Tani & Nolfi, 1999) or parametric bias shift (Tani, 2003) occurring in a step-wise fashion through the effort of minimizing prediction error, which drove the segmentation or raw perceptual flow into primitive sequences or chunks. After the learning process, the higher level network was also able to predict the sequences of behavior primitives in terms of shifts in this parametric bias vector. Tani interpreted this phenomena as the process of achieving a self-referential self, because the subjective experience of sensorimotor flow becomes objectified into reusable units which are manipulable by higher level processes, e.g., thinking. This interpretation is intuitive, because as the original experience of one's own sensorimotor flow is reconstructed with compositional structures, they become consciously describable objects rather than merely impressions of the original experiences. Then, from this understanding, Tani (2009) found in this capacity the origins of "self-referential selves" as the agent sets out actionable compositions as neurodynamical self-constructs that "emerge … through self-organizing compositional mechanisms of assembling and de-assembling sensorimotor schemata of repeated experiences", revisionary processes which arise only "in critical conditions of sustaining conflictive interactions between the top-down subjective mind and the bottom-up sensorimotor reality."[80]

The central thesis driving this work has been that consciousness arises in the correction and modification of dynamic structures potentially spontaneously generated in the higher-level cortical area including the pre-frontal cortex (PFC) in biological models. And importantly, a "simulated brain" has not been required to test this thesis. Instead, a much simpler system is able to embody its own possible future situations in the form of a minimal self, and to reconfigure this future as existing projects are frustrated. Tani employs the image of the sandpile, stable with every grain until the last when it all collapses to describe the condition of a dynamic system at a critical point. What results from the sand pile is another sandpile, with potential energy released that was otherwise bound up with its arrangement. The difference between a human being and a sandpile is more or less leisure, metabolism above background, in short a capacity to construct its own order through action. With the complexity of the evolved biological system, we may begin looking at constituent sub-sandpiles and their arrangements, as we had begun with biological systems in the first section of this paper. The questions that remain are merely how many one embodies, in what arrangement, which triggers first and in which contexts.[81]

One pressing objection to qualifying any such system as conscious, especially of the hard family of most- and myth-conscious, is that artificial systems are too simple. Of course, simulations of cognition are necessarily less complex than the biological system monitoring them, as overly-complex simulations defeat the purpose of a simulation.[82] Though it is true that artificial agents in a laboratory *are simpler than* organic brains out functioning in the real world, in artificial consciousness studies the potential exists to isolate essential features with a resolution otherwise lost against the background real-world noise and corruption of the biological approach. These artificial systems may be more conscious, completely conscious, or demonstrate pure consciousness in a way that a biological system cannot, because there are so many facets of cognitive agency essential to living systems that need not be replicated in an artificially conscious system. And, due to the nature of artificial systems, the hurdle that is privileged access to subjectivity may be overcome with perfect information about the dynamic structure of a cognitive system ready at hand. This potential does not exist in the study of biological systems. This potential is afforded, however, by artificial systems as we shall see in greater detail in the next paper in this series.

Artificial conscious systems afford a privileged insight into the structure of cognitive agency and how such systems in their normal operations result in the feelings of being a self in the world, a feeling that meets our every internal self-reflection. These investigations are not limited to available biological models, and there is no risk of polluting the natural environment with genetically engineered creatures designed to represent certain modes of consciousness over others. That said, biological studies continue to inspire artificial systems. For example, Tani's MTRNN architecture[83] was inspired by fMRI studies on higher level areas including the PFC showing them important to abstract reasoning and the integration of sensation. Research in mirror neurons inspired the hypothesis that predictive coding might be essential for pairing generation and recognition of actions, as tested with Tani's RNNPB.[84] However, the point is that empirical biological results cannot access the core problem of consciousness, as we have tried to articulate in the current paper.

CONCLUSIONS: WHAT TO EXPECT

The next paper in this series details how dynamic complex systems embodied in neurorobots demonstrate consciousness in their normal operations. The third paper in this series will revisit Boltuc's h-, and this paper's most- and myth- consciousnesses in order to evaluate Tani and colleagues' model. Are these robots h-conscious? More? Myth-conscious? At that point, finally we will revisit Boltuc's naturalistic non-reductionist thesis, as it may not be the material nature of a cognitive agent which ultimately grounds any account of consciousness, but rather the dynamic structure that had traditionally only existed in biological, and that now is instantiated also in artificial, forms.

ACKNOWLEDGEMENTS

Special thanks to Peter Boltuc for patient review and deft editing of multiple drafts, as well as to anonymous reviewers of this journal for recommending that we convert one long paper into a series over the course of the year.

NOTES

1. La Mettrie et al., *Man, a Machine: Including Frederick the Great's "Eulogy" on La Mettrie and Extracts from La Mettrie's "The Natural History of the Soul*," 98–99.
2. Boltuc, "The Philosophical Issue in Machine Consciousness," 159.
3. Ibid., 155.
4. Ibid.
5. Ibid., 160.
6. Ibid., 162. In example, he refers to the LIDA model, which posits a cycle of information processing within a Global Workspace, cf. page 157. We will briefly attend to the notion of the "global workspace" in the next section, but generally steer clear of the mire that is competing and complementary accounts of consciousness in the contemporary literature.
7. Ibid., 174.
8. Cf. Boltuc, 160.
9. Ibid., 162.
10. Ibid., 161.
11. Ibid.
12. Ibid., 158.
13. Ibid., 162.
14. Cf. Boltuc, 158.
15. Ibid., 173.
16. Ibid., 174.
17. See discussion in ibid., 174.
18. Bitbol, "On the Primary Nature of Consciousness," 268. See also White, *Conscience: Toward the Mechanism of Morality*; "Understanding and Augmenting Human Morality: An Introduction to the ACTWith Model of Conscience"; "Models of Moral Cognition," which articulate a similar integrative structure in terms of "stitching one's self into the world," and Chalmers, "How Can We Construct a Science of Consciousness?" for extended discussion of problems facing any such account.
19. Boltuc, "The Philosophical Issue in Machine Consciousness," 171.
20. Chalmers, *The Conscious Mind*.
21. The first author is indebted to Alexander VonSchoenborn for educating him to identify these IOUs.
22. Remaining sensitive to the fact that Boltuc's view on this point has developed in the meantime, the current paper attends solely to the position represented in his paper from 2009.
23. Cf. O'Muircheartaigh et al., "White Matter Connectivity of the Thalamus Delineates the Functional Architecture of Competing Thalamocortical Systems."
24. Cf. Llinas et al., "The Neuronal Basis for Consciousness." It is important to note that the systematicity of connected networks is reflected in the neural physiology of non-human animals, as well. See, for example, discussion in O'Muircheartaigh et al., "White Matter Connectivity of the Thalamus Delineates the Functional Architecture of Competing Thalamocortical Systems."
25. Parvizi and Damasio, "Consciousness and the Brainstem," 137.
26. Cf. Boltuc, "The Two Forks in the Road Towards h-consciousness," in press.
27. Bechara, Damasio, and Damasio, "Emotion, Decision Making, and the Orbitofrontal Cortex."
28. Cf. Coppin et al., "When Flexibility Is Stable: Implicit Long-Term Shaping of Olfactory Preferences."
29. Cf. Bechara, Tranel, and Damasio, "Characterization of the Decision-Making Deficit of Patients with Ventromedial Prefrontal Cortex Lesions," 2198.
30. Mitchell et al., "Medial Prefrontal Cortex Predicts Intertemporal Choice," 6.
31. Cf. Janowski, Camerer, and Rangel "Empathic Choice Involves vmPFC Value Signals That Are Modulated by Social Processing Implemented in IPL."
32. Cf. Koenigs et al., "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements."
33. See Mendez, "The Neurobiology of Moral Behavior: Review and Neuropsychiatric Implications," for review.
34. Cf. Damasio, *Self Comes to Mind: Constructing the Conscious Brain*.
35. Cf. Heidegger, *Being and Time: A Translation of Sein und Zeit*, re: "mitda-sein."
36. Cf. Spunt et al., "The Default Mode of Human Brain Function Primes the Intentional Stance."
37. The notion of most-consciousness is much more complicated than Boltuc's original h-consciousness. However, it better reflects the biological reality. Questions remain whether this "thickening" of the analytic sense of h-consciousness is a biological accident or essential to the dynamic structure necessary for cognitive agency in any form. And there is a role for analysis in answering these questions going forward. However, the purpose of analysis is to "carve the world at its joints," rendering complex systems simple enough for ready manipulation, and problems arise when analysis takes inquiry away from the original problem and directs instead to entities that exist only in the context of the analysis. One way to spot these problems is to put the parts together and see if anything is left over or left out. This is the method that we are pursuing in this series of papers, moving from biological to artificial systems integrations. And that said, one dimension of temporal integration remains left out of this model of most-consciousness.
38. Note that I do not write "generates within me the propositional attitude toward" (cf. Thagard, "Desires Are Not Propositional Attitudes").
39. Fuchs ("Self Across Time: The Diachronic Unity of Bodily Existence") recalls these sorts of phenomena as Leibniz's "little perceptions."
40. Note that Boltuc anticipates the form of this discussion, recognizing that there are in any conscious system also "unconscious cognitive processes" on the basis of which "one can ask whether an organism is conscious while still performing

apparent cognitive functions" (Boltuc, "The Philosophical Issue in Machine Consciousness," 160).

41. Again, we must here note that this portrait of h-consciousness is much richer than that of Boltuc's analysis. The view at root here can be found in White, "Models of Moral Cognition."

42. Cf. White, "What Happened? Alcohol, Memory Blackouts, and the Brain," for review.

43. Indeed, the alcohol industry might be characterized as an industry bent on these ends, to some degree, and if zombies were the object, then its potential in this effort is without question.

44. As much as there is a "problem of access" to another consciousness besides one's own, there is the reverse, an obligation to make one's own inner world sensible to others similarly embodied. Where this is impossible, agency is judged differently.

45. As understood, for example, in Habermas, *Communication and the Evolution of Society*.

46. The adolescence of the West after the death of the parents and the mourning period of the mid-life (as if life was a brutal and filthy orphanage), the collective anticipation to get out of that situation arose in the Enlightenment.

47. Boltuc, p. 160

48. Ibid. Note correlate portrait by Watanabe and Mizunami, "Pavlov's Cockroach: Classical Conditioning of Salivation in an Insect."

49. Edelman, Gally, and Baars, "Biology of Consciousness," 5.

50. For instance, implicating the hippocampus in Baars, Franklin, and Ramsoy, "Global Workspace Dynamics: Cortical 'Binding and Propagation' Enables Conscious Contents."

51. Or we may accept that it is only a feature of biological systems.

52. Edelman, Gally, and Baars, "Biology of Consciousness," 2.

53. Fuchs, "Self Across Time."

54. Ibid., 13.

55. Ibid., 8.

56. Ibid., 22.

57. Ibid., 15–16.

58. "We can define the entirety of established habits and skills as implicit or body memory that become current through the medium of the lived body without the need to remember earlier situations"(ibid., 16).

59. Ibid.

60. Cf. Zahavi, "The Experiential Self: Objections and Clarifications."

61. Fuchs, "Self Across Time," 8.

62. Ibid., 22.

63. Edelman, Baars, and Seth, "Identifying Hallmarks of Consciousness in Non-Mammalian Species."

64. It points to something essential so far neglected in our discussion. Take, for example, Aristotle's ideal life of study, ultimately centering on eternal relations and patterns displayed by increasingly perfect beings. Imagine that this is your life and that you wake into its confines. Little is pressing. Life itself is subsumed under an eternal order and your mind works to track this order, rather than remain sensitive to the minutia of daily psycho-political life. Now imagine that you are in an opposite condition, and whatever order on which you had depended in the world has disintegrated in short order. For the cosmological agent whose society may extend to the stars themselves in philosophical reflection on patterns of movement which also extend far beyond the scope of living anticipation, the sense that myth-consciousness extending beyond its embodied constraints (mine, yours, temporal, historical) is the sense into which one would awake. For the terrified agent, there is no anticipation of a future beyond immediate threats and immediately available resources. There is most-consciousness instead.

65. And for an essentially social being with a long early phase of direct mirroring of caregiver action and affect, followed only later with an equally long revisionary period alongside higher level neural growth and entrainment, most habits derive from those witnessed in others more or less alike. There remain questions about how these libraries are collated and how language emerges in distinct ways, permitting resonance between the similarly embodied while remaining exclusive of others. These questions can be answered within the context of artificial consciousness studies while remaining practically outside of any biological inquiry. For organisms like human beings, experience is historical-material of necessity. Where this is missing, there is deficiency. And due to deep material memory furnishing future self-situations extending well beyond any that an organism may itself potentially inhabit, for example, a just world after a few generations of adjustment for a childless man today, each moment is located within variable horizons of anticipation of succeeding moments against retained past, and not only for one's self and those alike, but potentially for any consciousness at all.

66. Fuchs, "Self Across Time," 12.

67. Ibid., 16.

68. Ibid., 17; quoting from Applefield, *The Story of a Life: A Memoir*, 90.

69. Jovasevic et al., "GABAergic Mechanisms Regulated by miR-33 Encode State-Dependent Fear."

70. Fuchs, "Self Across Time," 18.

71. See Abitbol et al., "Neural Mechanisms Underlying Contextual Dependency of Subjective Values: Converging Evidence from Monkeys and Humans," for review of common biological mechanisms.

72. Fuchs expresses doubts that "processes of vital self-regulation in the brainstem and diencephalon" may be replicated in an artificial agent, not solely due to their own complexity but because the brain is embedded in "rather slow biochemical interactions" which "may not be described as digital information processing in analogy to a computer" (14). In an artificial agent, these processes must take on a different form, and to which degree they need to be replicated remains to be seen.

73. Edelman, Gally, and Baars, "Biology of Consciousness," 5.

74. Implications of differing material modes of embodiment as this self is constituted are weighed in the third paper of this series.

75. Edelman, Gally, and Baars, "Biology of Consciousness," 5.

76. Cf. Tani and Nolfi, "Learning to Perceive the World as Articulated: An Approach for Hierarchical Learning in Sensory-Motor Systems"; Tani, "The Dynamical Systems Accounts for Phenomenology of Immanent Time: An Interpretation by Revisiting a Robotics Synthetic Study"; and Tani, *Exploring robotic minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena*

77. We set aside these issues for the next paper in order to quickly describe how a minimal self arises as a result of the basic interactions between these two embodied temporalities, and thus how any such system may appear as a "diachronic unity" in the first place.

78. Bak et al., "Self-Organized Criticality: An Explanation of the 1/fnoise."

79. Tani and Nolfi, "Learning to Perceive the World as Articulated: An Approach for Hierarchical Learning in Sensory-Motor Systems"; Tani, "Learning to Generate Articulated Behavior Through the Bottom-Up and the Top-Down Interaction Process."

80. Tani, "Autonomy of 'Self' at Criticality: The Perspective from Synthetic Neuro-Robotics," 423.

81. What is it like to be a sandpile.

82. Cf. White, "Simulation, Self-Extinction, and Philosophy in the Service of Science."

83. Yamashita and Tani, "Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: A Humanoid Robot Experiment."

84. Tani, "Learning to Generate Articulated Behavior Through the Bottom-Up and the Top-Down Interaction Process."

REFERENCES

Abitbol, R., M. Lebreton, G. Hollard, B. J. Richmond, S. Bouret, and M. Pessiglione. "Neural Mechanisms Underlying Contextual Dependency of Subjective Values: Converging Evidence from Monkeys and Humans." *Journal of Neuroscience* 35, no. 5 (2015): 2308–20.

Appelfeld, A. *The Story of a Life: A Memoir*. New York: Random House, 2009.

Bak, P., C. Tang, and K. Wiesenfeld. "Self-Organized Criticality: An Explanation of the 1/fnoise." *Physical Review Letters* 59, no. 4 (1987): 381–84.

Baars, B. J. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press, 1997.

Baars, B. J., S. Franklin, and T. Z. Ramsoy. "Global Workspace Dynamics: Cortical 'Binding and Propagation' Enables Conscious Contents." *Frontiers in Psychology* 4 (2013): 1–22.

Bechara, A., H. Damasio, and A. R. Damasio. "Emotion, Decision Making, and the Orbitofrontal Cortex." *Cerebral Cortex* 10, no. 3 (2000): 295–307.

Bechara, A., D. Tranel, and H. Damasio. "Characterization of the Decision-Making Deficit of Patients with Ventromedial Prefrontal Cortex Lesions." *Brain : a Journal of Neurology* 123 (2000): 2189–202.

Bitbol, M. "On the Primary Nature of Consciousness." In *The Systems View of Life*, edited by F. Capra and P. L. Luisi, 266–68. Cambridge: Cambridge University Press, 2014.

Boltuc, P. "The Philosophical Issue in Machine Consciousness." *International Journal of Machine Consciousness* 1, no. 1 (2009): 155–76.

Boltuc, P. "The Two Forks in the Road Towards h-consciousness." *Theoria et Historia Scientiarum* (in press, 2016).

Chalmers, D. J. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press, 1996.

Chalmers, D. J. "How Can We Construct a Science of Consciousness?" *Annals of the New York Academy of Sciences* 1303, no. 1 (2013): 25–35.

Coppin, G., S. Delplanque, C. Porcherot, I. Cayeux, and D. Sander. "When Flexibility Is Stable: Implicit Long-Term Shaping of Olfactory Preferences." *PLoS ONE* 7, no. 6 (2012): e37857. http://doi.org/10.1371/journal.pone.0037857

Damasio, A. R. *Self Comes to Mind: Constructing the Conscious Brain*. New York: Pantheon Books, 2010.

Edelman, D. B., B. J. Baars, and A. K. Seth. "Identifying Hallmarks of Consciousness in Non-Mammalian Species." *Consciousness and Cognition* 14, no. 1 (2005): 169–87.

Edelman, G. M., J. A. Gally, and B. J. Baars. "Biology of Consciousness." *Frontiers in Psychology* 2, no. 4 (2011): 1–7. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3111444

Fuchs, T. "Self Across Time: The Diachronic Unity of Bodily Existence." *Phenomenology and the Cognitive Sciences* (forthcoming, 2016) doi: 10.1007/s11097-015-9449-4

Habermas, J. *Communication and the Evolution of Society*. Boston: Beacon Press, 1979.

Heidegger, M., and J. Stambaugh. *Being and Time: A Translation of Sein und Zeit*. Albany, NY: State University of New York Press, 1996.

Janowski, V., C. Camerer, and A. Rangel. "Empathic Choice Involves vmPFC Value Signals That Are Modulated by Social Processing Implemented in IPL." *Social Cognitive and Affective Neuroscience* 8, no. 2 (2013): 201–08.

Jovasevic, V., K. A. Corcoran, K. Leaderbrand, N. Yamawaki, A. L. Guedea, H. J. Chen, G. M. G. Shepherd, and J. Radulovic. "GABAergic Mechanisms Regulated by miR-33 Encode State-Dependent Fear." *Nature Neuroscience* 18, no. 9 (2015): 1265–71.

Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio. "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgements." *Nature* 446, no. 7138 (2007): 908–11.

La Mettrie, J. O., G. C. Bussey, M. W. Calkins, and Frederick, King of Prussia. *Man, a Machine: Including Frederick the Great's "Eulogy" on La Mettrie and Extracts from La Mettrie's "The Natural History of the Soul."* Chicago, Illinois: Open Court Publishing Co., 1912.

Llinas, R., U. Ribary, D. Contreras, and C. Pedroarena. "The Neuronal Basis for Consciousness." *Philosophical Transactions: Biological Sciences* 353, no. 1377 (1998): 1841–49.

Mendez, M. F. "The Neurobiology of Moral Behavior: Review and Neuropsychiatric Implications." *CNS Spectrums* 14, no. 11 (2009): 608–20. Last accessed January 31, 2016, at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3163302/

Mitchell, J. P., J. Schirmer, D. L. Ames, and D. T. Gilbert. "Medial Prefrontal Cortex Predicts Intertemporal Choice." *Journal of Cognitive Neuroscience* 23, no. 4 (2011): 857–66.

O'Muircheartaigh, J., S. S. Keller, G. J. Barker, and M. P. Richardson. M. P. "White Matter Connectivity of the Thalamus Delineates the Functional Architecture of Competing Thalamocortical Systems." *Cerebral Cortex* 25, no. 11 (2015): 4477–89.

Parvizi, J., and A. Damasio. "Consciousness and the Brainstem." *Cognition* 79 (2001): 135–59.

Spunt, R. P., M. L. Meyer, and M. D. Lieberman. "The Default Mode of Human Brain Function Primes the Intentional Stance." *Journal of Cognitive Neuroscience* 27, no. 6 (2015): 1116–24.

Tani, J. "An Interpretation of the 'Self' from the Dynamical Systems Perspective: A Constructivist Approach." *Journal of Consciousness Studies* 5, nos. 5-6 (1998): 516–42.

Tani, J., and S. Nolfi. "Learning to Perceive the World as Articulated: An Approach for Hierarchical Learning in Sensory-Motor Systems." *Neural Networks*, 12, no. 7 (1999): 1131–41.

Tani, J. "Learning to Generate Articulated Behavior Through the Bottom-Up and the Top-Down Interaction Process." *Neural Networks* 16 (2003): 11–23.

Tani, J., and M. Ito. "Self-Organization of Behavioral Primitives as Multiple Attractor Dynamics: A Robot Experiment." *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* 33, no. 4 (2003): 481–88.

Tani, J. "The Dynamical Systems Accounts for Phenomenology of Immanent Time: An Interpretation by Revisiting a Robotics Synthetic Study." *Journal of Consciousness Studies* 11, no. 9 (2004): 5–24.

Tani, J. "Autonomy of 'Self' at Criticality: The Perspective from Synthetic Neuro-Robotics." *Adaptive Behavior* 17, no. 5 (2009): 421–43.

Tani, J. *Exploring robotic minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena.* New York: Oxford University Press, forthcoming.

Thagard, P. "Desires Are Not Propositional Attitudes." *Dialogue: Canadian Philosophical Review / Revue Canadienne De Philosophie* 45, no. 1 (2006): 151–56.

Watanabe, H., and M. Mizunami. "Pavlov's Cockroach: Classical Conditioning of Salivation in an Insect." *PLoS ONE*, 2, no. 6 (2007): e529. doi:10.1371/journal.pone.0000529

White, A. M. "What Happened? Alcohol, Memory Blackouts, and the Brain." *Alcohol Research & Health: The Journal of the National Institute on Alcohol Abuse and Alcoholism* 27, no. 2 (2003): 186–96.

White, J. B. *Conscience: Toward the Mechanism of Morality*. Columbia, MO: University of Missouri–Columbia, 2006.

White, J. "Understanding and Augmenting Human Morality: An Introduction to the ACTWith Model of Conscience." *Studies in Computational Intelligence* 314 (2010): 607–21.

White, J. "Models of Moral Cognition." In *Model-Based Reasoning in Science and Technology: Theoretical and Cognitive Issues*, edited by L. Magnani, 363–91. Berlin: Springer, 2014.

White, J. "Simulation, Self-Extinction, and Philosophy in the Service of Science." *AI & Society* (2015) doi:10.1007/s00146-015-0620-9

Yamashita, Y., and J. Tani. "Emergence of Functional Hierarchy in a Multiple Timescale Neural Network Model: A Humanoid Robot Experiment." *PLoS Computational Biology* 4, no. 11 (2008): e1000220.

Zahavi, D. "The Experiential Self: Objections and Clarifications." In *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*, edited by M. Siderits, E. Thompson, and D. Zahavi. Oxford: Oxford University Press, 2010.