Design for a Darwinian Brain:
Part 1. Philosophy and Neuroscience

Chrisantha Fernando

Dept. of Electronic Engineering and Computer Science
Queen Mary University of London
ctf20@eecs.qmul.ac.uk

Abstract. Fodor and Pylyshyn in their 1988 paper denounced the claims of the connectionists, claims that continue to percolate through neuroscience. In they proposed that a physical symbol system was necessary for open-ended cognition. What is a physical symbol system, and how can one be implemented in the brain? A way to understand them is by comparison of thought to chemistry. Both have systematicity, productivity and compositionality, elements lacking in most computational neuroscience models. To remedy this woeful situation, I examine cognitive architectures capable of open-ended cognition, and think how to implement them in a neuronal substrate. I motivate a cognitive architecture that evolves physical symbol systems in the brain. In Part 2 of this paper pair develops this architecture and proposes a possible neuronal implementation.

1   Introduction

This is the first of a two-part paper showing that physical symbol system with the properties defined by Fodor and Pylyshyn could plausibly exist in the brain. I will explain what a physical symbol system is and what properties to expect from it, by showing how biochemical systems in cells can be interpreted as physical symbol systems. Having described the substrate independent properties of a physical symbol system, I will then describe two possible implementations of physical symbol systems in the brain, one based on activity patterns, and the other based on connectivity patterns. Finally I consider cognitive architectures that are physical symbol systems, and discuss their current algorithmic limitations, in order to motivate our own architecture described in Part 2.

The Hungarian chemist Tibor Ganti described the concept of fluid automata, or chemical machines [1] that are made of molecules. Molecules are objects composed of atoms that have specific structural relations to other atoms in the molecule. A molecule is assembled according to a combinatorial syntax, i.e. a set of chemical structural constraints such as valance, charge, etc that determine how atoms can legally join to make a molecule. Combinatorial semantics determines how a molecule with a particular structure will react (behave) in a given environment. Semantic content in terms of a chemical symbol system equates

to its chemical function, i.e. its reactivity. The function of a molecule is a function of the semantic content of its parts, translated; the reactivity of a benzene ring is modified by its side-groups such as methyl groups. In short, the chemical symbols and their structural properties cause the system behaviour.

This physical symbol system composed of a vat of chemical molecules operates in parallel rather than in series. It is constrained by kinetics. Its function (reactivity) is subjected to non-encoded influences such as temperature that influences the thermodynamic equilibrium position of chemical reactions. These aspects do not tend to intrude in normal conversations about physical symbol systems, but they are germane to processing taking place in physical symbol systems made of chemicals. A pleasing example of a symbolically specified computation in chemistry is a chemical clock. The two autocatalytic cycles of the Belousov-Zhabotinsky (BZ) reaction constitute a fluid automaton that implements a chemical clock. Whilst it is the symbolic organisation of the molecules that specifies the reaction network topology, it is at the level of the analog (dynamical systems) operations of the assembled reaction network that the clock like phenomenon of the BZ reaction exists. Continuous behavior results from the ensemble properties of a multitude of discrete symbolic chemical operations. The tremendous success of chemistry is to have robustly linked the chemical system level with the macroscopic level.

The fact the chemical molecules are physical symbol systems gives the macroscopic properties of chemistry the same abstract characteristics found in any physical system system. These are now examined. For example, chemistry has productivity. Productivity means that a system can encode indefinitely many propositions, that there is recursive assembly of parts of the representation to produce the proposition, and that an unbounded number of non-atomic representations is entailed [2]. If we replace the term proposition with adaptation, then in biological evolution, the number of possible adaptations is unlimited in the same sense, being the developmental result of a symbolic genetic and epigenetic encoding. Whether it is helpful to replace proposition with adaptation is a moot point, but to me this seems natural because adaptations have a form of truth-value in the sense that they are the entities to which fitness applies. Similarly, the capacity for chemical reactivity is unlimited, i.e. there are many more possible chemical reactions than could be implemented in any realistically sized system. An unlimited number of molecules can be produced allowing an unlimited many chemical reactions, and this is made possible with only a finite set of distinct atom types. Therefore, an unbounded set of chemical structures must be composite molecules. In physical symbol systems terms it is also the case that an indefinite number of propositions can be entertained, or sentences spoken. This is known as the productivity of thought and language. If neural symbol systems exist, they must have the capacity of being combined in unlimited ways. The next section deals with how this could potentially work. Finally, there is reason to believe that no non-human animal has the capacity for productive thought [3].

The second property of chemistry is that it is systematic, i.e. the capacity for atoms to be combined in certain ways to produce some molecules is intrinsically connected to their ability to produce others. This is neatly shown by how chemists learn chemistry. There are several heuristics that chemists can learn to help predict properties of molecules and how they will react based on their structure. Only a chemist lacking in insight will attempt to rote learn reactions. Similarly with systematicity in language. The ability to produce and understand a sentence is intrinsically connected with the ability to produce and understand other sentences. Languages are not learned by learning a phrasebook, they have syntax which allows the systematic combination of entities.

The third property of chemistry is that an atom of the same type makes roughly the same contribution to each molecule in which it occurs. For example, the contribution of a hydrogen atom to a water molecule influences many reactivity properties of that molecule. Hydrogen atoms suck electrons and this is a property of the hydrogen atom itself, whatever in most cases it binds to. This means there is systematicity in reactivity (semantics) as well as in structure (syntax). This is known as compositionality in physical symbol system speak. In the same way, lexical items in sentences have approximately the same contribution to each expression in which they occur. This approximate nature suggests that there is a more fundamental set of atoms in language than words themselves, e.g. linguistic constructions.

Why was the idea of a physical symbol system even entertained in chemistry? Why did people come to believe in discrete atoms coming together systematically to form molecules, instead of for example some strange distributed connectionist model of chemistry? Lavoisier discovered a systematic relationship between chemical reactions, i.e. the conservation of mass. Proust discovered the law of definite proportions, i.e. that compounds can be broken down to produce constituents in fixed proportions. Dalton extended this law of multiple proportions that explained that when two elements come together to form different compounds (notably the oxides of metals), they would come together in different small integer proportions. Are there any analogous reasons to believe in physical symbol systems in the brain, based on observation of human thought and language?

Open-ended cognition appears to require the ability to learn a new symbol system, not just to implement an already evolved one. Children can learn and manipulate explicit rules [4] and this implies the existence of a neuronal symbol system capable of forming structured representations and learning rules for operating on those representations. For example, Gary Marcus has shown that 7 month old infants can distinguish between sound patterns of the form ABA versus ABB, where A and B can consist of different sounds e.g. "foo". "baa" etc. Crucially, these children can generalize this discrimination capacity to new sounds that they have never heard before, as long as they are of the form ABA or ABB. Marcus claims that performance in this task requires that the child must extract "abstract algebra-like rules that represent relationships between placeholders (variables), such as the first item X is the same as the third item Y,

or more generally that item I is the same as item J" Several attempts have been made to explain the performance of these children without a PSS (e.g. using connectionist models) but Marcus has criticised these as smuggling in symbolic rules in one way or another by design. For Marcus it seems that the system itself must discover the general rule. In summary, the problem with a large set of connectionist learning devices is that a regularity learned in one component of the solution representation is not applied/generalized effectively to another part. Marcus calls this the problem of training independence, a solution to which is one of the fundamental requirements for a learning system to be described as symbolic or rule based.

Concerning the discussion on whether connectionist models without explicit symbols nor rules could be sufficient for open-ended cognition it is informative to quote the discussion of limitations of their own connectionist/computational neuroscience model given by Rougier et al (2005) [5]. Their model was able to form representations of categories of class in the Wisconsin Card Sorting Task. The authors in the main paper claim to produce complex representations without symbols. However, they moderate their claims in the Supplementary Material as follows. "Our model does not directly address this capacity for generativity... it was never asked to perform entirely new tasks... people can recombine familiar behaviors in novel ways. Such recombination, when built upon a rich vocabulary of primitives, may support a broad range of behavior and thus contribute to generatively.". The paper acknowledges the simplicity of the algorithm implemented, i.e. "random sampling + delayed replacement" and acknowledges the need for mechanisms that can undertake more sophisticated forms of search. "Generativity also highlights the centrality of search processes to find and activate the appropriate combination of representations for a given task, a point that has long been recognized by symbolic models of cognition... Our model implemented a relatively simple form of search based on stabilizing and destabilizing PFC representations as a function of task performance, which amounts to random sampling with delayed replacement... An important focus of future research will be to identify neural mechanisms that implement more sophisticated forms of search.".

Another paper of interest is that of Tani and Nolfi [6]. It self-organizes the connections between neural module experts on the basis of minimising of prediction error of the expert predictors at the lower levels. The higher level then predicts the gate opening dynamics of the lower level predictors and propagates this upwards. Two levels are modelled. It is claimed that such systems are an alternative to explicit symbolic models of complex environments. However, the system is very severely theoretically limited. The reason is that for a high dimensional space of models, there would be a combinatorial explosion of possible higher-order models which could be composed from combinations of lower order models. This problem has also been faced by Goran Gordon in a set of papers about hierarchical curiosity loops [7]. A generative process for constructing novel experts is needed. Each module effectively then becomes an atom in a physical symbol system, with its inputs and outputs specifying its connections to other

atoms, i.e. permitting arbitrary variable binding. Incidentally, the use of prediction error also becomes infeasible for open-ended systems because the space of trivial predictions is also too large. Therefore, neither of these two connectionist papers propose how complex representations can emerge without explicit symbols. The former because of reasons the authors admit themselves, and the later due to its non-extensibility to high dimensional systems.

In summary, the following is a concise definition of a symbol system adapted from Harnad to emphasise the chemical aspects [8]. A symbol system contains a set of arbitrary atoms (or physical tokens) that are manipulated on the basis of explicit rules that are likewise physical tokens or strings (or more complex structures, e.g. graphs or trees) consisting of such physical tokens. The explicit rules of chemistry generate reactions from the structure of atoms and molecules (plus some implicit effects, e.g. temperature). The rule-governed symbol-token manipulation is based purely on the shape of the symbol tokens (not their meaning), i.e., it is purely syntactic, and consists of rulefully combining and recombining symbol tokens, in chemical reactions. There are primitive atomic symbol tokens and composite symbol-token strings (molecules). The entire system and all its parts – the atomic tokens, the composite tokens, the syntactic manipulations both actual and possible and the rules - are all "semantically interpretable". The syntax can be systematically assigned a meaning e.g., as standing for objects or as describing states of affairs. For example, semantic interpretation in chemistry means that the chemical system exhibits chemical reactivity. In biochemistry this is extended to a higher level. In the chemicals inside cells, it is fair to say that properties of those chemicals, e.g. the concentration, or their configuration, actually stands for states of affairs in the environment outside the cell, for example the conformation of a signalling molecule may represent the glucose concentration outside the cell. In the same way a neural symbol system exhibits behavior such as the child's capacity to distinguish ABA from ABB in grammar learning tasks.

This analogy has helped me to understand what Fodor and Pylyshyn were talking about in their famous paper denouncing connectionism [2]. From this basis I developed what I believe may be two plausible candidates for a neuronal symbol system. The first has been described in detail in two previous publications (Fernando 2011; Fernando 2011). In short it consists of arbitrary physical tokens (spatiotemporal patterns of spikes) arranged into molecules or symbol structures (groups of spatiotemporal patterns of spikes passing through the same network). These tokens undergo explicit rule-governed symbol-token manipulation (reactions involving transformation of the spike patterns), and a process of stochastic optimization can learn the explicit rule sets. Thus the proposed neural representation of an atomic symbol token is a spatiotemporal patterns of spikes. Neurons detect a particular spatiotemporal spike pattern if the axonal delays from the pre-synaptic neuron to the detector neuron are properly matched to the spatiotemporal pattern such that depolarization reaches the detector neuron body simultaneously. This implementation of neural symbol-tokens (atoms) uses the concept of polychronous computing and a modification of the

concept of wavefront computing [9] [10]. The construction of molecular symbol structures from atomic symbol-tokens requires binding of atomic symbol-tokens together such that they can be subsequently manipulated (reacted) as a function of the structure of the molecule. In my framework, compositional neural symbolic structures exist as temporally ordered sequences of symbols along chains of neurons. A symbol-structure can be described as a sequential combination of spike patterns, a feature that has been described in real neuronal networks as a "cortical song" [11]. A great many such short chains of combined spatiotemporal spike patterns may exist in the brain. Each chain being akin to a register in a computer, blackboard or tape that can store symbol-tokens of the appropriate size. A single symbol-token could be read by a detector neuron with the appropriate axonal delay pattern when interfacing with the chain. A parallel symbol system in the brain consisting of a population of such chains, each capable of storing a set of symbol token molecules and operating on these molecules in parallel is conceivable. Interaction between (and within) such chains constitutes the operations of symbol-manipulation. Returning to the chemical metaphor, such interactions can be thought of as chemical reactions between molecules contained on separate chains, and rearrangements within a molecule expressed on the same chain. Whilst in a sense a chain can be thought of as a tape in a Turing machine (due to the serial nature of the strings), it also can be thought of as a single molecule in a chemical system (due to the existence of multiple parallel chains). This constitutes the core representational substrate on which symbol manipulation will act.

Receiving input from the chain and writing activity back into the chain is done by a detector neuron with specific input and output delays in relation to the chain. A detector neuron only fires when the correct pattern of input is detected. In effect the neuron is a classifier. Where the classifier fails to fire, the same pattern enters the chain as leaves the chain. This is because the spatiotemporal organization of these patterns does not match the spatiotemporal tuning curve of the detector neuron. Only when the spatiotemporal spike pattern matches does the detector neuron fire. Once fired, the output of the detector neuron is injected back to the neurons of the chain. If the output of the detector neuron slightly precedes the normal passage of the untransformed pattern through the chain, then the refractory period of the output neurons of the chain prevents interference by the original untransformed pattern, which is thereby replaced by the new pattern specified by the detector neuron. Such a detector neuron we will now call a classifier neuron because it is a simple context free re-write rule with a condition (detection) and an action pole of the type seen in Learning Classifier Systems (LCS) [12] [13].

It can be seen that such classifier neurons are selective filters, i.e. the classifier neuron is only activated if the spatiotemporal pattern is sufficiently matched with the axonal delays afferent upon the neuron. The above classifier implements an implicit rule. An implicit rule is a rule that operates on atomic or molecular symbol structures without being specified (encoded/determined/controlled) by a symbol structure itself. There is no way that a change in the symbol system,

i.e. the set of symbols in the population of chains, could modify this implicit matching rule. The implicit rule is specified external to the symbol system. Whenever the symbol passes along this chain, it will be replaced by the new symbol, irrespective of the presence of other symbols in the system.

In a symbol system (as in chemistry), symbols are manipulated (partly) on the basis of explicit rules . This means that the operations or reactivity of symbols depends on/is controlled by/is causally influenced by their syntactic and semantic relationship to other symbols within the symbol-structure and between symbol structures. If inhibitory gating neurons are combined with classifier neurons then it becomes possible to implement explicit rules within our framework. If an inhibitory gating unit must receive a particular pattern of spikes in order for it to become active, and if that unit disinhibits a classifier neuron that is sensitive to another symbol token then it is possible to implement a context-sensitive re-write rule. The rule is called context sensitive because the conversion of X to Y depends on the relation of X to another contextual symbol T. A set of context-sensitive re-write rules is capable of generating a grammar of spike-patterns. Consider starting the system off with a single symbol-token S. Probabalistic application of the rules to the initial symbol S would result in the systematic production of spike patterns consisting of grammatically correct context-sensitive spike pattern based sentences. A major implementation issue in real neuronal tissue would be the fidelity of transmission of spatiotemporal spike patterns. The information capacity of such a channel may fall off with decreasing fidelity of copying in that channel in a manner analogous to Eigen's error catastrophe in genetic evolution [14]. Finally, we have shown how it is possible to undertake stochastic optimization in the above system by copying with mutation of neuronal classifiers [15]. This mechanism depends on spike-time dependent plasiticity (STDP) by which one neuron infers the re-write rule implemented by another neuron, with errors. This permits a population of re-write rules to be evolved. The details of this mechanism are beyond the scope of this paper but can be found elsewhere [15].

What kinds of behavior require a physical symbol system, if any? We have understood something about how chemical and genetic systems generate reactions and cellular behavior. But what about how neural representations produce organismal behavior? Whilst a gene is a functional unit of evolution, we can ask, what are the possible functional units of thought at the algorithmic level, i.e. what are the entities in the brain that encode behavioral traits whose probability of transmission during learning is correlated with reward (or value functions of reward) [16] [17]. It may well be the case that there exist units of thought that do not have compositionality, systematicity and productivity at all. For example, a solitary feed-forward neural network that implements a forward model which predicts the next state of the finger as a function of the previous joint angles and the motor command given to the finger, does not satisfy the definition of a symbol system discussed above. Such models of the world may be complex, e.g. inverse models of the motor commands required to reach a desired goal state, but may lack the properties of a physical symbol system, e.g. [6] and [5]. The units may contribute to many domains from the linguistic to the sensorimotor,

or occupy a level of specialized niches, e.g. units in the cerebellum, or units in the primary visual cortex. By analogy with units of evolution [18] they may be called units of thought, although their individuality may be more a convenience for measurement and description than a strict reality, for example it is possible that they are physically overlapping. I define them as the entities in the brain whose survival depends upon value, whatever form such value takes, e.g. extrinsic or intrinsic values [19]. How far can an animal get in cognition without physical symbol systems? In short the mystery is this; what are the kinds of representation upon which search must act in order to produce cognition, and how do certain classes of representation map to certain classes of cognition? A concrete proposal is given in Part 2. The short answer is that it is open-ended generative processes that require physical symbol systems, of the type for example that Rougier et al say their connectionist model cannot do [5].

The second question is what kinds of learning algorithm are good for generative search? In answering this question I have found it important to notice that both thought and evolution are intelligent, and undertake knowledge-based search as defined by Newell [20]. I propose that both open-ended thought and open-ended evolution share similar mechanisms and that a close examination of these two processes in context is helpful. Traditionally thought has been considered to be directed and rational [21], whereas evolution by natural selection has been considered random and blind [22]. Neither view in their extreme forms is correct. A necessary feature of creative thought is stochastic search [23] [24] and evolution by natural selection whilst it may be a blind watchmaker is not a stupid one, it can learn to direct its guesses about what variants are likely to be fit in the next generation [25]. Apparently a grandmaster, maybe Reti, was once asked "How many moves do you look ahead in chess?" to which he answered"One, the right one!" We now know that evolution makes moves (mutants) in the same way, not by explicitly looking ahead, but by recognizing what kinds of moves tended to be good in the past in certain situations and biasing mutation accordingly [26] [27][28]. In fact, this principle has been exploited in the most sophisticated black-box optimization algorithms such as covariance matrix adaptation evolution strategies (CMA-ES) that rivals the most sophisticated reinforcement learning algorithms [29]. How do the mechanisms of cognition differ from the mechanisms of genetics in discovering how to make new adaptive ideas/organisms effectively? It seems that processes of systematic inference are in place in both adaptive systems, e.g. bacterial genomes are organized into Operons that through linkage disequilibrium tend to bias variability between generations along certain dimensions, i.e. crossover is a structure sensitive operation, a notion that was coined to describe the Classical cognitive paradigm [2]. Structure sensitive operations are the symbolic equivalent of methods to bias variation in continuous systems such as CMA-ES.

Both evolution and thought are generative processes with systematicity and compositionality because both depend on structure sensitive operations on representations [2] that far exceed the complexity of CMA-ES. We are able to generate educated guesses about hidden causes, to make up explanations for which there

is as yet insufficient data to use deduction. This is evidenced by children's ability to learn language [30][31][32]. In fact there is considerable evidence much of which I find convincing, that no non-human animal can entertain models of the world that require the hypothesis of unlimited kinds of hidden causes [3]. Genetic and epigenetic generative mechanisms seem rather inflexible in comparison to those of thought. From close examination of these abilities can we have some idea what units of thought must be capable of representing about the world, i.e. how they limit the kinds of concept we can entertain? Perhaps we have a chance by looking at an example of constrained creativity from Gestalt psychology in the form of insight problems [33]. The 9-dot problem is a task in which you must cover each dot in a 3 x 3 matrix of dots with just 4 straight lines without removing your pen from the paper. Through the application of introspection we seem to solve this problem by generating a set of possible solutions, inventing some kind of fitness function for the solutions such as how many dots they cover, and mutating the better solutions preferentially in a systematic way. We may not be able to verbalize the scoring function nor how the initial set of solutions was chosen or varied. Often we will reach an impasse, a kind of sub-optimal equilibrium in thinking about this problem. The impasse may be diffused by suddenly discovering the correct solution that arises in a punctuated event called the Ah-ha moment. How are the solutions to the above problem represented, how many distinct solutions are represented at once in the brain, how are they varied over time, how do solutions interact with each other, how does a population of solutions determine behavior, and how are the scoring functions represented in the brain, and how many there are? We have proposed elsewhere that the solution of such problems involves a Darwinian search in the brain [34]. It may be the case that careful modelling of the above kinds of thinking process can uncover the kinds of representation and the search algorithm used to search in the space of such representations.

In cognitive science it is useful to put your money where your mouth is and actually try to build a robot to do something. Part 2 shows how we propose to build a robot that has open-ended creativity and curiosity that exhibits productivity, compositionality, and systematicity of thought. The most advanced attempts to build robots capable of open-ended creativity are briefly discussed. The general principle in all that work has been to learn predictive models of self and world. These are typically implemented as supervised learning devices such as neural networks. The networks are not physical symbol systems. However, some approaches attempt to combine predictors compositionally. The approach that I have decided to take in the remainder of Part 1 is to examine the limitations of these approaches in producing productive behavior in robots.

The approach taken by most groups in trying to achieve open-ended cognition is to learn predictive models of the world that become as rich as possible, and to guide action selection on the basis of how well certain models are being learned. The idea is that as we grow up we try to model the world in greater and greater richness and depth. An early and beautiful approach to this problem was by A.M. Uttley who invented a conditional probability computer and connected

it to a Meccano robot that was controlled with a hard-wired Braitenberg type controller [35][36] . This computer learned to associate sensory states with motor actions when this external controller was controlling the robot, and could learn the sensorimotor rules, e.g. turn left if left sensor dark and right sensor light, and eventually substitute this hard-wired controller. Note that this was an entirely associationist form of model and would fall under what Fodor and Pylyshyn call Connectionism. More recent approaches learn state(t)-motor(t)-next state(t+1) triplets through experience, generating a network of anticipations [37]. Once such a cognitive map is learned it can be used for model-based reinforcement learning (dynamic programming) to backpropogate rewards associated with goal sensory states to earlier states and thus to permit planning of high-value trajectories. The capacity to simulate sense-action trajectories is a critical function of a cognitive map. The sophistication of mental simulation is a significant constraint on intelligence and creativity. Note that there is no physical symbol system involved in the above algorithm for cognitive map building. How then might a physical symbol system help in learning models of the world?

An obvious limitation of the above kind of cognitive map formation is that it may fail when there are unobserved causes. One solution is the generation of classifiers that infer the existence of hidden-causes [38] and then utilize this in further modelling of the world. By this means an unlimited conceptual world opens. This is where physical symbol systems may well enter the picture. Once a space of higher-level classifiers exists, the combinatorial explosion of possible conditions and predictions requires an effective method for sparse search in this higher-order space. The space of higher-order predictions is also highly structured. In summary, the units of thought in TGNG are s(t)-m(t)-s(t+1) triplets that store transition probabilities. There is no need for compositionality and systematicity of such units because they can be entirely instructed by supervised learning methods (at least in low dimensional sensorimotor systems). Efficient sparse search algorithms would however be needed when hypothesising hidden states for which there is insufficient sensory evidence to undertake supervised learning. So, by thinking about the problem from this robotic perspective, I have identified a domain in which stochastic processes and representations with compositionality and systematicity would be expected to bring efficiency, i.e. during search in a structured space of hidden causes.

Interestingly, a modification of the simple cognitive maps above that potentially adds compositionality and systematicity has been proposed in order to scale reinforcement learning to "domains involving a large space of possible world states or a large set of possible actions" [39]. The solution has been to chunk actions into higher-level actions consisting of sequences of primitive actions, called "options" [40]. Options can be hierarchically organized policies. Each abstract action specifies a policy (a map from states to actions) that once started from in initiation state, executes a sequence of actions (which may also be options) until a termination state is reached. They work by reducing the size of the effective search space by allowing structuring of search by modular recombination (concatenation) of options. The critical question is how the correct set of useful

options can be chosen in the first place. Typically options are hand-designed, with options specifying sub-goals to be achieved. In none of the models are options stochastically explored. The options framework is an admirable first step towards achieving a truly systematic and compositional physical symbol system for describing actions. It is related to the notion of action grammar developed by Patricia Greenfield (Greenfield 1991) which also seeks to identify the behavioural correlates of neural symbol systems.

To conclude, for any reasonable sized sensory and motor dimensionality, and for reasonable time-periods, exhaustive search in predictive model space is not possible. The computation is exponential in time and space. To solve this problem, units of thought (e.g. predictors and actors) should exhibit compositionality and systematicity, of the kind for example that we see in the options framework. The system should have the ability to form hierarchical descriptions of action and hierarchical predictive models. Part 2 presents our fluid option framework or fluid action grammar in which option-like symbol structures are co-evolved in the brain along with their fitness functions.

References

1. Ganti, T.: The Principles of Life. Oxford University Press, Oxford (2003)
2. Fodor, J., Pylyshyn, Z.: Connectionism and cognitive architecture: A critical analysis. Cognition 28, 3–71 (1988)
3. Penn, D., Holyoak, K., Povinelli, D.: Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. Behavioral and Brain Sciences 31(2), 109–130 (2008)
4. Marcus, G.: The Algebraic Mind: Integrating Connectionism and Cognitive Science. MIT Press (2001)
5. Rougier, N., Noelle, D., Braver, T., Cohen, D., O'Reilly, R.: Prefrontal cortex and flexible cognitive control: Rules without symbols. Proc. Natl. Acad. Sci. U. S. A. 102, 7338–7343 (2005)
6. Tani, J., Nolfi, S.: Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems. Neural Networks 12, 1131–1141 (1999)
7. Gordon, G.: Hierarchical exhaustive construction of autonomously learning networks (2011)
8. Harnad, S.: The symbol grounding problem. Physica D 42, 335–346 (1990)
9. Izhikevich, E.M.: Polychronization: computation with spikes. Neural Computation 18(2), 245–282 (2006)
10. Izhikevich, E.M., Hoppensteadt, F.: Polychronous wavefront computations. International Journal of Bifurcation and Chaos 19, 1733–1739 (2009)
11. Ikegaya, Y., et al.: Synfire chains and cortical songs: Temporal modules of cortical activity. Science 304, 559–564 (2004)
12. Holland, J., Reitman, J.: Cognitive systems based on adaptive algorithms. ACM SIGART Bulletin 63, 43–49 (1977)
13. Wilson, R.: Function approximation with a classifier system (2001)
14. Eigen, M.: Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58(10), 465–523 (1971)

15. Fernando, C.: Symbol manipulation and rule learning in spiking neuronal networks. Journal of Theoretical Biology 275, 29–41 (2011)
16. Price, G.: Selection and covariance. Nature 227, 520–521 (1970)
17. Price, G.: The nature of selection. Journal of Theoretical Biology 175(3), 389–396 (1995)
18. Maynard Smith, J.: The problems of biology. Oxford University Press, Oxford (1986)
19. Stout, A., Konidaris, G., Barto, A.: Intrinsically motivated reinforcement learning: A promising framework for developmental robotics (2005)
20. Newell, A.: Unified Theories of Cognition. Harvard University Press (1990)
21. Newell, A., Simon, H.A.: Human problem solving. Prentice-Hall, Englewood Cliffs (1972)
22. Dawkins, R.: The Selfish Gene. Oxford University Press, Oxford (1976)
23. Perkins, D.: Insight in minds and genes. In: Sternberg, R., Davidson, J. (eds.) The Nature of Insight, MIT Press, Cambridge (1995)
24. Simonton, D.: Foresight in insight? a darwinian answer. In: Sternberg, R., Davidson, J. (eds.) The Nature of Insight. MIT Press, Cambridge (1995)
25. Pigliucci, M.: Is evolvability evolvable? Nature Reviews Genetics 9, 75–82 (2008)
26. Kashtan, N., Alon, U.: Spontaneous evolution of modularity and network motifs. Proc. Natl. Acad. Sci. U. S. A. 102(39), 13773–13778 (2005)
27. Izquierdo, E., Fernando, C.: The evolution of evolvability in gene transcription networks (2008)
28. Parter, M., Kashtan, N., Alon, U.: Facilitated variation: How evolution learns from past environments to generalize to new environments. PLoS Computational Biology 4(11), e1000206 (2008)
29. Stulp, F., Sigaud, O.: Path integral policy improvement with covariance matrix adaptation (2012)
30. Goldberg, A.: Constructions: A Construction Grammar Approach to Argument Structure. University of Chicago Press., Chicago (1995)
31. Steels, L.: Experiments on the emergence of human communication. Trends in Cognitive Sciences 10(8), 347–349 (2006)
32. Steels, L., De Beule, J.: Unify and merge in fluid construction grammar (2006)
33. Sternberg, R., Davidson, J.: The Nature of Insight. MIT Press, Cambridge (1995)
34. Szathmary, E., Fernando, C.: Concluding remarks. In: The Major Transitions Revisited, MIT Press, Cambridge (2009)
35. Andrew, A.: Machines that learn. The New Scientist, 1388–1391 (1958)
36. Uttley, A.: Conditional probability computing in a nervous system. In: Mechanisation of Thought Processes, vol. 1, pp. 121–147. H.M. Stationery Office, London (1959)
37. Butz, M., Shirinov, E., Reif, K.: Self-organizing sensorimotor maps plus internal motivations yield animal-like behavior. Adaptive Behavior 18(3-4) (2010)
38. Nessler, B., Pfeiffer, M., Maass, M.: Stdp enables spiking neurons to detect hidden causes of their inputs (2010)
39. Botvinick, M., Niv, Y., Barto, A.: Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. Cognition 113(3), 262–280 (2009)
40. Sutton, R., Precup, D., Singh, S.: Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. Artificial Intelligence 112, 181–211 (1999)