By using this website, you consent to the use of cookies as described in our Privacy Policy

Advances in Artificial Intelligence
Volume 2013, Article ID 841646, 13 pages
http://dx.doi.org/10.1155/2013/841646

Research Article

Selection for Reinforcement-Free Learning Ability as an Organizing Factor in the Evolution of Cognition

Solvi Arnold, Reiji Suzuki, and Takaya Arita

Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Received 28 August 2012; Revised 21 January 2013; Accepted 5 February 2013

Academic Editor: Bikramjit Banerjee

Abstract

This research explores the relation between environmental structure and neurocognitive structure. We hypothesize that selection pressure on abilities for efficient learning (especially in settings with limited or no reward information) translates into selection pressure on correspondence relations between neurocognitive and environmental structure, since such correspondence allows for simple changes in the environment to be handled with simple learning updates in neurocognitive structure. We present a model in which a simple form of reinforcement-free learning is evolved in neural networks using neuromodulation and analyze the effect this selection for learning ability has on the virtual species' neural organization. We find a higher degree of organization than in a control population evolved without learning ability and discuss the relation between the observed neural structure and the environmental structure. We discuss our findings in the context of the environmental complexity thesis, the Baldwin effect, and other interactions between adaptation processes.

1. Introduction

This paper explores the relation between the structure of an environment and the structure of cognitions evolved in that environment. Intuitively, one would expect a strong relation between the two. In the past, some have taken this intuition very far. Spencer [1] viewed the evolution of life and mind as a process of internalization of progressively more intricate and abstract features of the environment. He traced the acquisition of such "correspondence" between the internal and external from basic life processes (e.g., the shape of an enzyme molecule has a direct and physical relation to the shape of the molecule whose reactions it evolved to catalyze), all the way up to cognitive processes (such as acquisition of complex causal relations between entities removed in space and time). That a certain correspondence should exist between the shapes of enzyme and substrate will be uncontroversial, but how far can this concept of correspondence take us when cognition is concerned?

Certainly, when we hand-code an AI to function within a given environment, we can typically recognize much of the environmental organization in the structure of our AIs' cognitions. However, as the history of connectionism demonstrates, fit behaviour does not necessarily involve intelligible neural structure. More often than not, the neural organization of evolved artificial neural networks (ANNs) allows little if any interpretation in terms of environmental structure. If we demand that models of the mind in some sense "reflect" their environment, then the lack of Spencerian correspondence in evolved ANNs poses a conundrum. One possible response is to abandon our "correspondence intuition" altogether [2]. Another response is to declare such ANNs unfit as models of cognition (see, e.g., [3–5]). We believe that both of these responses in fact mask a deeper issue: we do not actually have a clear understanding of how cognitive evolution arrives at the sort of clearly structured solutions that we find in ourselves (and observe in numerous other species).

For exploring this issue, ANNs should be an excellent tool, since they allow for large variation in what we might call their "degree of organization". By evolving networks that initially lack organization in various environments and under various constraints, we can study the processes that give them shape and structure and identify the environmental features that drive those processes. This type of approach is found in work on the evolution of modularity, for example [6–8]. Here we apply this sort of approach to studying the effect of learning on emergence of neurocognitive organization.

Let us first have a closer look at the intuitions that our messy ANNs seem to violate. Our own cognition seems quite organized: we have a high degree of functional differentiation, with distinct innate cognitive faculties specializing on distinct aspects of the environment. In a sense, the

environment we evolved is reflected in our innate cognitive and neural architecture. On a shorter timescale, our thoughts and mental representations refer to and individuate the contents of the world around us. Advanced cognition does not *just work somehow* (as is often said of ANNs); it works by actively establishing (both on evolutionary and lifetime timescales) highly specific correspondence relations with the environment.

In stark contrast, connectionism forcefully reminds us of the unintuitive fact that any mapping that can be implemented can be implemented in infinitely many ways. Implementations composed of building blocks mapping one to one to the building blocks of the environment are but one option, and there is a whole spectrum of viable candidates seemingly reaching from one to one across many to many all the way to all to all. Why should evolution prefer implementations that allow interpretation in terms of Spencerian one to one correspondence? Connectionism shows us that, *in general, it doesn't*. Evolution is a behaviourist. If your behaviour is fit, then you are fit, regardless of the details of your cognitive architecture. But we can ask a better question: under what circumstances *does* evolution favour correspondence? Building on our previous work [9], we present a hypothesis on how the evolution of learning ability leads evolution to correspondence, and we provide a proof of principle in the form of a simple artificial life model in which we simulate the evolution of a simple form of reinforcement-free learning. Section 2 explains the theory, Section 3 explains the choice for reinforcement-free learning, Section 4 describes our computational model, and results are analyzed in Section 5, followed by discussion and conclusions in Sections 6 and 7.

2. Learning and Neurocognitive Organization

That the evolution of learning should somehow lead to correspondence may seem like an odd suggestion: the history of connectionism has made it quite clear that regardless of whether we train them (using learning algorithms like e.g. error back-propagation) or evolve them (using genetic algorithms), artificial neural network structures usually do not admit much interpretation in terms of the environments they were trained or evolved in. Why would a combination of learning and evolution do any better? The answer lies in how these two adaptation processes interact (Figure 1).

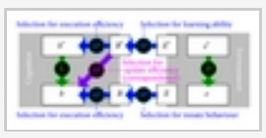

Figure 1: Selection for correspondence. SP: selection pressures/evolutionary adaptation processes. L: learning/lifetime adaptation processes (implementation of learning ability updates implementation of behaviour). D: environmental dynamics. Selection for learning implies selection for update efficiency in the implementation of behaviour, that is, alignment of the dynamics of B with the dynamics of E. This alignment depends on correspondence between the structure of the behaviour implementation and the structure of the environment, leading to selection on correspondence (diagonal arrow).

We define *behaviour* as a mapping from stimuli to responses:

$$B : S \longrightarrow R. \tag{1}$$

We define learning as a mapping from stimulus-behaviour pairs to behaviours, that is, "stimulus-caused updates of behaviour":

$$\begin{aligned} B &: (S, B) \longrightarrow B = \\ B' &: (S, S \longrightarrow R) \longrightarrow (S \longrightarrow R). \end{aligned} \tag{2}$$

Learning updates behaviour on a within-lifetime timescale and both behaviour and learning are products of evolution. We also define the environment as a mapping, one from responses to stimuli. An agent acts, and this (potentially) affects the subsequent stimulus it receives:

$$E : R \longrightarrow S. \tag{3}$$

Note that an environment is much like an inverse behaviour (mapping responses to stimuli instead of stimuli to responses). Just as behaviour may change, so may the environment, either as the result of the agent's responses or spontaneously. We denote change in E as E' without specifying it further.

We have defined behaviour, learning, and environment as mappings, but organisms and environments are physical objects, not mathematical objects. In order for these mappings to exist in the physical world they must have implementations. For each of the mappings defined above, we let its lowercase partner denote its implementation: b, b', e, e'. Here e and e' should be understood as the actual physical reality of the environment. In reality, e and e' are generally not clearly distinguishable. Physically speaking, learning occurs via modification of b by b'.

Selection for a given behaviour (B) is selection for implementation of that behaviour (b), but as noted before, any given behaviour can be implemented in infinitely many ways. We may expect evolution to favour implementations that execute efficiently (in terms of time or energy consumption of whatever is precious in a given setting), but as connectionist history shows, selection for execution efficiency should not be expected to produce correspondence.

If the environment is dynamic in a sufficiently organized and predictable way as to make learning possible, then there is selection pressure on evolution of B' and hence b'. Analogously to B and b above, execution efficiency in b' may be selected for but this should not be expected to produce correspondence. But something interesting happens between b' and b. Given that b' must update b, different b call for different b'. For example in the highly unnatural case that b would take the form of a table defining an output for each possible input independently, then b' would operate by rewriting entries of this table. So whether and how feasible evolution of B' is strongly depends on the architecture of b. If there is selection pressure on B', then mutations in b that are beneficial to B' are beneficial mutations (even if they have no effect whatsoever on B). As an extreme scenario, we could imagine B remaining stable while b evolves to facilitate B'. This possibility shows that there is a fundamental difference between selection for a specific mapping and selection for a specific implementation of that mapping.

So while evolution working on B alone does not care much about the structure of b, "coevolution" (if we may abuse the term a little) of B' and b *does* care about the structure of b. Along the long horizontal arrows in Figure 1, evolution treats its objects as black boxes (selecting on input-output relations alone, i.e., B and B'), but indirectly, via selection pressure on learning, it peeks inside b and selects for implementation structure (diagonal arrow).

B' constrains b, but we have not said anything yet about what sort of b is favoured by B'. We will claim that B' benefits most from b that employs correspondence with the environment. The basic idea is as follows: if the environment and (consequently) the optimal behaviour are static, then

difference in the structure of their implementations poses no problem. But if the environment and (consequently) the optimal behaviour may change (by means of E' and B', respectively), then the more the structure of b and e differ, the harder it is for B' to update B in sync with E'. The implementations (e) of environments that cognition evolves in are composed of distinct aspects (food sources, temperatures, other agents, spatial layouts, etc.) that act and interact to give rise to E. Let us call a change in one such aspect a *simple* change. Simple changes in e often lead to *complex* changes in E: multiple input-output pairs change. Consequently a complex update of B is required. If b contains an aspect corresponding to the changed aspect of e, in a functionally similar position, then the required complex change in B can be realized by a simple change in b. This makes B' quite feasible. If no such corresponding aspect exists, a complex implementation update is required. In this case no straight-forward relation exists between the environmental change and the appropriate behaviour change, making B''s work difficult or infeasible.

So the organization that evolves in b to facilitate B' should in one form or another capture the variable aspects of the environment along with their functional roles therein. This is what we mean by correspondence, and also what we take Spencer to mean by correspondence. Note that we do neither claim that B' is strictly impossible without correspondence between e and b nor that such correspondence cannot occur in absence of B'. What we claim is that selection pressure on B' translates into selection pressure on correspondence between e and b, and that this "selection pressure conversion" is an organizing factor in the evolution of cognition.

Note that, in general, not all of b receives this organizing influence. Innate behaviour that is impervious to modification by learning should not be affected. As such the hypothesis here recuperates neither our intuitions nor Spencer's correspondence-based theory of the evolution of mind in full. However, the parts of b that *are* modifiable by learning seem quite central to advanced cognition, and the hypothesis provides a candidate explanation of why these parts should be as organized as they are. One especially notable aspect of cognition that (in our present conceptualization) falls outside the scope of this organizing influence is learning ability itself (as it is not affected by learning). Learning ability could be placed under similar organizing influence by introducing a second-order learning process updating the first (see discussion), but we do not further consider this possibility here.

Our concept of correspondence here is intentionally broad, as there may be a lot of variation in how correspondence might be realized. However, reasoning from the purpose served by correspondence, we can make some more easily objectively verifiable predictions. If our reasoning is correct, then behaviour implementations evolved under selection for learning should distinguish themselves from implementations under no such pressure by (1) functional differentiation and (2) compactness.

(1) Functional differentiation: if distinct aspects of the environment are in some sense replicated in the implementation of behaviour, then we should expect to find functionally distinct substructures in that implementation.
(2) Compactness: if no updating of behaviour is required, then nothing keeps the implementation from spreading out over the available implementation substance. If updating is required, then correspondence helps to minimize the amount of physical change that needs to be made to realize the required updates. This should constrain distribution and promote a more focused implementation in which large, controlled behaviour updates can be made with minimal physical change.

In assessing our results we will consider these measures in addition to our assessment of correspondence.

Before we move on to the next section, let us briefly address a complicating factor in establishing correspondence. We discussed how, given correspondence between b and e, simple change in e could be met with simple change in b. It could be objected that, more often than not, changes in e are not simple. Consider the changing of the seasons. Although triggered by a simple change in a distal aspect of e (the angle at which the earth faces the sun), the effect on the local environment is a complex change (many aspects of e change), which reveals itself to cognitions in that environment as a complex change in E. Given their common cause, the simple changes comprising the complex change in e will cooccur, and we may find it being handled with a simple change in some aspect of b. Can this aspect of b now be said to correspond? If so, what to? To all the aspects of e that changed or to the aspect that triggered all those other changes (the angle at which the earth faces the sun)? In the former case we would have a one to many correspondence, which strays from our original concept somewhat, and in the latter case we would have a correspondence with something so distal that calling it a correspondence seems odd. We will not attempt to answer this or other questions arising from causal relations between environmental changes here. Regardless of whether and what to this aspect of b would correspond, we can see that having it would be advantageous precisely in environments with changing seasons, so organization is still selected for by environmental dynamics.

3. Reinforcement-Free Learning

If we hypothesize internalization of environmental regularities to be of importance for efficient learning, then the question arises why traditional machine learning algorithms (supervised learning and reinforcement learning) seem to do fine without. We think the answer is that such algorithms model only some very limited subclasses of the learning abilities found in nature. While undemanding in terms of internalization of environmental regularities, they require explicit examples or rewards to drive the learning process. When examples or rewards are explicitly provided, then all that is needed to make learning possible is an ability to recognize examples/rewards as such (and this ability is tacitly assumed in such learning algorithms). (This is enough to make learning *possible*, but in most cases such learning could be made more efficient by exploitation of environmental regularities, as was shown in [9].) We will not discuss example-driven learning here, as its limitations are evident. However when we look at reward-based learning, there is a relevant parallel between the fields of machine learning and psychology.

Reward-based learning algorithms are designed to be applicable without prior knowledge of the environment. Behaviourism tried to capture all learning in terms of universally applicable conditioning rules. In this sense traditional machine learning and behaviourism are alike: they deal in universal, reinforcement-driven learning rules. In both cases, one could say that it is the aim for universality that prevents them from reaching an accurate understanding of learning as it occurs in nature, as most if not all of the learning ability found in nature is nonuniversal and based on more or other information than a simple reward signal (the clearest example being language acquisition). Such nonuniversal learning, sensitive to and capable of using the information its specific environment provides, is where dependence on environmental structure, and hence selection for correspondence, can be expected to be most prominent.

So how can we computationally approach nonuniversal learning ability as it occurs in nature? We believe that the most sensible approach is to evolve learning ability from scratch, starting with a mechanism for *arbitrary behaviour change*, and letting it evolve into forms of learning ability that will use not just reward information but whatever available information it can find a use for. Examples of learning evolved from mechanisms for behaviour change are found in [10–15]. Performance wise, such systems do not yet compare favourably to conventional machine learning approaches. However, if the aim is to study evolution of cognition as it occurs in nature, we believe this approach to be the correct one.

In the present paper, we intentionally omit a reward-signal altogether, forcing evolution of *reinforcement-free* learning ability. The motivation for this choice is twofold: it eliminates the risk of dependence on reward muddying our results and shows how evolved learning can solve a task that would leave standard reinforcement learning algorithms stuck at zero. The next section explains our computational approach.

4. Evolving Reinforcement-Free Learning

If, as argued in the previous section, selection pressure on learning ability translates into selection pressure on alignment of cognitive and environmental organization, then it should be possible to make such alignment evolve by evolving learning ability. We start with a description of our baseline model, in which behaviour is shaped by evolution alone. In this model, no learning occurs (in the terminology of Section 2, we evolve b in absence of selection pressure on B'). Then we extend this baseline model (introducing selection for learning and a mechanism for neural plasticity) to create our main model in which the same behaviour as in the baseline model is acquired via a combination of evolution and learning instead. We compare the networks evolved in these two models to see whether they show any differences that pertain to our hypothesis.

4.1. Baseline Model

A population of simple feed-forward neural networks was evolved to "catch" prey in a simple toroidal 20×20 binary grid world. The agent's body occupies one cell. Prey are represented by 2 adjacent 1-cells. To avoid ambiguity, two prey objects never occupy adjacent cells (e.g., a 2×2 block of 1-cells could equally well represent two vertically oriented prey or two horizontally oriented prey, so we prevent such ambiguous configurations from occurring). The environment always contains 40 prey. Interaction with prey is illustrated in Figure 2. Prey is "caught" by stepping on it from a suitable angle, yielding one fitness point. Stepping on it from a bad angle instead costs one fitness point. In either case, the prey disappears and respawns outside the net's field of view. Prey are otherwise stationary. The species has a repertoire of 4 movement actions: step forward, jump forward, turn right, and turn left. Turns are always 90 degrees. The jump action moves the organism forward by two cells instead of one, without touching the cell in between (so this action can be used to jump over prey when the step action would cause incorrect approach, and it is also generally the faster mode of movement). Each individual has its own private environment, so no interaction between individuals occurs.

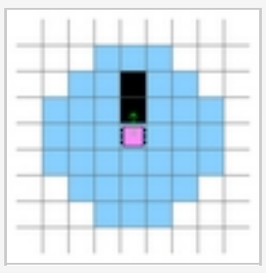

Figure 2: Catching prey. Up is forward. The turquoise area indicates the field of view. (a) If the prey is stepped or jumped on longitudinally, the organism gains one fitness point (the prey is caught and eaten). (b) If the prey is stepped or jumped on laterally, the organism loses one fitness points (the prey injures the organism and flees).

The neural networks take as input the organism's field of view (turquoise area in Figure 2), taking the cell values (0 or 1) as activation values for the input neurons. The input layer additionally contains a bias neuron with its activation always set to 1.0, and a noise neuron with its activation set to a random value in [0, 1] at every time step (included to provide a source of randomness to drive exploratory behaviour, if such behaviour were to be selected for), making for a total of 39 input neurons. There is a hidden layer of 16 neurons, using the hyperbolic tangent activation function (activation range $[-1, +1]$). There are 4 output neurons, one for each action. At each time step, the action of the output neuron with the highest activation value is performed.

Connection weights are evolved using a simple Genetic Algorithm. 10 independent trials of the baseline model were performed (i.e., 10 independent populations of networks were evolved), using the following settings: lifetime: 400 time steps, population size: 90, parent pool size: 40, elite group size: 10 (parent pool and elite group are nonexclusive). Individuals in the elite group are copied unaltered to the next generation; individuals in the parent group produce two offspring each, to which mutation is applied. Mutation is single point. Selection is rank based. To get a good spread of heavily and subtly mutated individuals, a genome to be mutated first gets a random mutation sensitivity value in the range [0.0, 0.0075]. This sensitivity value is then used as the mutation probability for every connection weight in the network. Mutation of a connection has a 0.75 probability of adding a random value from $[-0.5, +0.5]$ to the connection weight and a 0.25 probability of resetting the connection to 0. Weights are clipped to the $[-2, +2]$ range, and connection weights of the first generation are also randomly picked from this range.

Unsurprisingly, the population quickly evolved the ability to approach prey correctly and scored high fitness ever after (Figure 3). In the next section we use these scores as a baseline for assessing the efficacy of learning ability evolved in the main model.

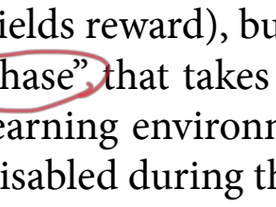

Figure 3: Evolution process of baseline model. Averages over 10 runs. Gray areas indicate standard deviation over the runs. Note the log scale on the x-axis.

4.2. Main Model

For the main model, we introduce an element of unpredictability in order to create selection pressure on learning ability, and we introduce a mechanism for behaviour change (*not* a learning algorithm) that can be *evolved into* the necessary learning ability.

The unpredictability we use is randomization of the assignment of actions to output neurons. In the baseline model, this assignment is invariant over the population (e.g., for any individual, output neuron #1 was assigned the "step" action, etc.). In the main model, the assignment is randomized for every new individual. Consequently, every individual first has to learn how to control itself before it can competently catch prey. In our terminology of Section 2, this variation in output-action assignment is the environmental dynamic E' that necessitates evolution of B', which should lead to incorporation of structural features of e into b.

This learning process could be driven by reward (the organism could experiment with various actions in various states and reinforce whatever yields reward), but as our aim is to evolve reinforcement-free learning ability, we explicitly block this possibility as follows: we introduce a "learning phase" that takes place in a "learning environment", in which there is sensory input but no reward. Individuals first spend 800 time steps in this learning environment and are then placed in the regular environment where their performance is assessed as in the baseline model. Learning is disabled during this performance phase, to guarantee that all learning takes place in the learning environment.

In the learning environment there are no prey objects, but "toy" objects, which consist of a single 1-cell (to avoid ambiguity, toy objects never occupy adjacent cells). The learning environment is otherwise identical to the performance environment. Toy objects have no effect on fitness, but they provide sensory feedback when an organism performs an action. For example, when an organism takes a step forward, the positions of all objects within the field of view will be shifted down by one cell. Each action is associated with such a characteristic transformation of the content of the field of view, so the environment provides the information necessary to overcome the randomization of the action-output assignment, but without offering any reinforcement to drive a reward-based learning process.

Note that there is no target behaviour for the learning phase (since no fitness is awarded, any behaviour is as good as any other). Moreover, the states that will provide opportunities to score rewards never occur during the learning phase (as there are no prey objects in the learning environment).

Here we evolve this unusual type of learning ability, using a modification of the neuromodulation concept of Soltoggio et al. [14, 15]. The idea is to include special "modulatory" neurons in the network, that can dynamically alter the plasticity of the connections of the neurons they project to. These modulatory neurons mostly behave like regular neurons (they take input from regular neurons and apply their activation function like regular neurons), but instead of an activation signal they send out a modulation signal. Regular neurons sum the modulation signals they receive, apply their activation function, and then use the resulting value as their plasticity value for the present time step. This dynamic plasticity is then used as learning rate in an otherwise Hebbian weight update. This update rule captures the core concept of heterosynaptic plasticity as is known to underpin much learning ability in biological brains, albeit in highly simplified and abstracted form (biological neuromodulation comprises a collection of complex chemical processes modifying the way neural activation affects synapse strength).

For the main model we added 6 of these modulatory neurons in the hidden layer of the network species used in the baseline model (Figure 4). These too use the hyperbolic tangent activation function. Since individuals must learn by observing how their actions transform the content of their field of view, some form of memory of the previous time step is necessary. We provide this memory by adding "echo neurons" to the input layer. These trivially copy the pattern of the regular input neurons, but with one time step of lag. To keep comparison with the baseline model fair, the echo neurons are only connected to the modulatory neurons (so they only serve learning and do not directly influence action selection). Since weight updates only occur during the learning phase, neither the modulatory neurons nor the echo neurons have any influence whatsoever during the performance phase. In the terminology of Section 2, the input, hidden, and output neurons comprise b, while the echo and modulatory neurons comprise b'. The update rule for connection weights is as follows:

$$\Delta W_{ij} = 0.01 \cdot m_i \cdot a_i \cdot a_j, \tag{4}$$

where W_{ij} is the weight of the connection from neuron i to neuron j, m_i is the modulation at neuron i, and a_i is the activation of neuron i. Weights are again clipped to the range $[-2, +2]$. The update rule contains a multiplication by 0.01 to avoid excessively large sudden weight changes. Our experiments suggest that a wide range of values can be used here. No attempts were made to optimize this value. Weights are updated after observation of the state resulting from the chosen action. Note that since only the hidden neurons receive modulation, only the top layer connections are updated during the lifetime.

Figure 4: Network topology used in main model. Light-blue arrows are regular activatory connections (full connectivity). The purple arrow indicates modulatory connections (full connectivity). The dashed arrow indicates a copying operation with one step of lag (one to one connectivity, see text). The input, hidden and output neurons occur in both models; the echo and modulatory neurons are unique to the main model.

We provide the learning process information on which action was performed using what can be thought of as mutual inhibition between output neurons: the activation of the output neuron whose action is performed is set to 1, and the activation of the other output neurons is set to 0 (mutual inhibition was also applied in the baseline model but serves no function there).

We found that for connections exposed to learning, very small innate weights work best. These are hard to evolve with the mutation settings of the baseline model, so for these connections mutation strength and clipping range are divided by 200. All other connections use the same parameters as in the baseline model.

The baseline model provides an upper limit for what the main model might achieve. The performance of the main model should approach that of the baseline model to the extent that the learning ability evolved in the main model can overcome the randomization of the action-output assignment. (The comparison we make might seem odd, since we varied two aspects (task and network type) instead of one. However, varying just one of these makes little sense: evolving nets without learning ability in an environment that requires it is not informative (average fitness remains stuck at around 5), and evolving nets with learning ability in an environment that does not require it is unlikely to produce interesting learning ability. The aspect we aim to vary is how behaviour is realized: by evolution alone or by evolution and learning together. Varying this aspect requires varying both task and network type.) Figure 5 shows performance of the main model. Evolution in the main model is slower than in the baseline model, as is to be expected given the increased complexity of the task and the larger genotype. Once stabilized, generation best and generation average fitness values do not quite match those of the baseline model, but they get fairly close (±109 versus ±100 and ±87 versus ±65, resp.). The difference in genotype size accounts for part of the difference in average performance (both models use the same range of per-connection mutation probabilities, and the networks in the main model have substantially more connections, so we should expect to find a larger proportion of dysfunctional mutants in the population). For comparison, evolving populations without learning ability under the randomization condition of the main model produces average fitness values of about 5 (results not shown). Considering that the main model's nets have just an 800-time-step learning phase to wire up a solution from experience while the baseline model's nets benefit from 50000 generation of connection weight refinement, we can conclude that the evolved learning ability copes with the output-action assignment randomization fairly well.

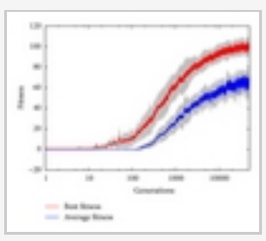

Figure 5: Evolution process of main model. Averages over 10 runs. Gray areas indicate standard deviation over the runs. Note the log scale on the x-axis.

5. Network Structure Analysis

The results discussed so far show that it is quite possible to evolve reinforcement-free learning ability using a neuromodulation approach. Next we discuss the effect of the need for learning on the neural organization of the networks, assessing whether or not evolution of learning ability leads to increased organization, and whether their organization expresses any aspects of the environment.

Figures 7 and 8 show the connection patterns of the hidden neurons of the best individual of the last generation of each run for baseline and main model, respectively. The plot format is explained in Figure 6. Looking at Figures 7 and 8 we can note two differences. (1) In the main model, we see that many hidden neurons end up with uniform upward connections (visible in that the four columns showing upward connections of such neurons all take on the same colour). Given that actions are selected by picking the output neuron with the highest activation, we can conclude that these neurons are not functionally involved in action selection. (Since this observation could possibly be explainable as an effect of the relatively small number of modulatory neurons, we performed a small number of runs with 12 instead of 6 modulatory neurons (results not shown). These did not reveal a higher number of functionally involved neurons, suggesting that a small number of functionally involved neurons (i.e., a compact solution) is somehow advantageous.) Neurons like these are rare or absent in the baseline model. (2) In the main model, we see a small number of connection patterns occur time and time again over multiple runs. While not absent in the baseline model, this tendency is notably stronger in the main model.

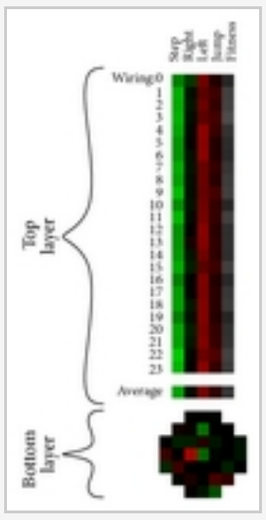

Figure 6: Explanation of neural connectivity plots used in Figures 7 and 8. (a) This plot shows the weights of all incoming and outgoing connections of a single hidden neuron. The bottom part shows weights of incoming connections (from input neurons), arranged after the positions in the field of view that the input neuron gets its input from (e.g., we see that this example neuron has a negative connection to the input neuron that perceives the state of the cell directly to the left of the organism and a positive connection to the input neuron that perceives the state of the cell two steps ahead of the organism). These weights are innately fixed in both models. In the main model, the outgoing connections (to output neurons, top part of plot) are subjected to learning. To plot these, we ran the individual with every possible output-action assignment (24 assignments in total), repeating each assignment 25 times. The connection weights shown (numbered rows) are averages over the 25 lifetimes per assignment, and a global average over all assignments. For convenience, we arrange the weights of the outgoing connections not by output but by action of the output neuron the connection projects to (e.g., the first column shows strength of connection to the output neuron assigned to the "step" action, regardless of which output neuron that is under a given output-action assignment). In the baseline model the weights of all connections are constant over the lifetime and only a single output-action assignment exists, so in plots of baseline model networks there is only a genetic weight to show for each outgoing connection. The rightmost column of the top part shows the fitness of the individual containing the neuron. In case of the main model, fitness is shown as average per assignment and as global average, like the connection weights. For the baseline model, we show a global average over the same number of lifetime runs used for the global averages of the main model, that is, $24 \times 25 = 600$. Note that fitness columns are identical for all neurons from a single individual. (b) Flipping the sign on all incoming and outgoing connections yields a functionally identical connection pattern. This symmetry was taken into account in further analysis. (c) Colour scales for connection weight and fitness value.

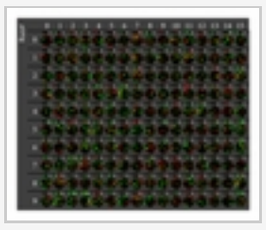

Figure 7: Connectivity of hidden neurons in best individual of final generation for 10 runs of the baseline model. Each row of 16 neurons represents one such individual (see Figure 6(a) for explanation of plot).

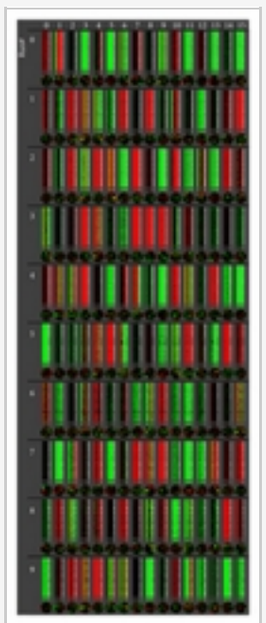

Figure 8: Connectivity of hidden neurons in best individual of final generation for 10 runs of the main model. Each row of 16 neurons represents one such individual (see Figure 6(a) for explanation of plot).

Figure 9 shows the result of CT clustering [16] (modified to handle sign symmetry; see Figure 6(b)) applied to the hidden neurons' connection patterns. We see that CT clustering identified a conspicuous trio of clusters of seven neurons each. This trio of clusters reflects a single solution that evolved in 7 out of the 10 runs of the main model. (2 runs evolved another 3-neuron solution. The remaining run was not stable at the time of termination and had traits of both solutions.) Although similar connections patterns do occur in the baseline model (most notably connection pattern #4 of the main model is common in the baseline model as well), the variability in connection patterns is larger there (resulting in a larger number of clusters), and the tendency to evolve characteristic sets of neurons is weaker. Figure 10 gives an idealization and explanation of the solution highlighted in Figure 9. We can easily see how this triplet produces the behaviour the environment demands.

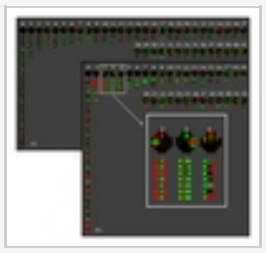

Figure 9: We pooled all neurons shown in Figure 8 and pooled all neurons shown in Figure 9, then applied CT clustering [16] to each pool in order to detect evolutionary trends over the different runs of the models. Clustering was performed on basis on the incoming connections only (as these are set by evolution only in both models). Each column shows one cluster of neurons. The pattern at the top of each column is an average over the downward connections of all neurons in the cluster. Below it are the upward connection patterns for each member neuron (for the main model, these are averages as shown in Figure 7(a)). Connection patterns were sign-flipped whenever this led to smaller distance to cluster center, to account for sign symmetry (see Figure 7(b)). (a) Result for baseline model. (b) Result for main model. The inset shows the triplet of neurons identified as the common solution evolved in 7 of the 10 runs of the main model. This triplet alone suffices for adequate behaviour.

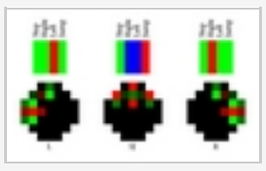

Figure 10: Idealization of the solution evolved in runs 0, 1, 2, 3, 5, 7, and 8 of the main model. These patterns were named L (for left turn), R (for right turn) and SJ (for step or Jump). Signs of the L pattern are flipped with respect to Figure 9. L and R detect horizontally oriented prey to the left and right, respectively, and promote the corresponding turn action when appropriate (by increasing activation on the output neuron assigned the “turn left" action and lowering activation on all other output neurons). When moving neither right nor left, SJ assesses whether it is better to step or jump (by increasing activation on the output neuron assigned the preferred action, and lowering activation on the other). Connections coloured blue varied over different occurrences of the SJ neuron (they do not affect behaviour).

How do these findings relate to our theory? We said that a species that needs to learn will tend to evolve solutions that are (1) more compact, (2) show increased functional differentiation, and (3) exploit environmental regularities. That the solutions evolved in the main model are more compact than those of the baseline model is visible in the larger number of neurons that is not functionally involved in action selection. As for functional differentiation, we might have expected to see neurons specializing on a single specific action or situation, but the results here are clearly different. We see that functionally involved neurons in the main model specialize not on single actions but on making specific distinctions required for correct action choice (caption Figure 10). This sharp role division between neurons, each handling a specific distinction, constitutes clear functional differentiation. The reason for specialization on distinctions becomes clear when we search for exploitation of environmental regularities.

The question whether the evolved neural networks exploit environmental regularities is more difficult to answer, but we believe the answer to be yes. Consider what happens during the learning process when an upward connection is established. When the L neuron of Figure 10 is negatively connected to the left-turn output neuron, then observation of horizontal prey to the left comes to promote leftward turns. Additionally, as a side effect, vertically oriented prey straight ahead will come to *inhibit* leftward turns. These two effects are not a priori related. That turning left is a bad idea when there is a prey straight ahead might seem trivial when phrased like this, but bear in mind that the spatial coherence of the environment is never explicitly given and that the output- action wiring is randomized. That this feature of the learning system is remarkable will be clearer when we phrase it from the nets’ point of view:

That action **a** *will favourably transform input vector* **P** *implies that action* **a** *will NOT favourably transform input vector* **Q**.

This relation evidently does not hold in general. There are plenty of different input vector configurations that are all advantageously transformed by the same action. Of course in a spatially coherent environment, there will be many pairs of input vectors for which the above relation *does* hold (if **P** is an input vector representing horizontally oriented prey to the left and **Q** is an input vector representing prey straight ahead, the above relation holds for **a** = “turn left”), but we cannot a priori identify them *without reference to that environment’s spatial coherence*. Similar arguments can be made for the R and SJ neurons.

Similar arguments seem also applicable for the alternative solution that evolved in 2 runs. Given this solution’s low number of occurrences we cannot extract as clear an image from the data, and the instances we have seem more diffuse than the instances of the more common solution, but Figure 11 attempts to illustrate the basic mechanism. We see that control is similarly carved up into a small number of choices, but the choices implemented by two of the neurons (TF and LR in Figure 11) differ from the more common solution.

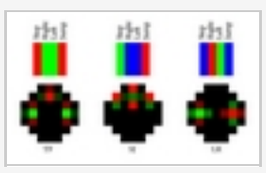

Figure 11: Attempted idealization of the solution evolved in runs 4 and 9 of the main model. These patterns were named TF (for turn or Forward), LR (for left or Right) and SJ (for step or Jump). TF promotes the step and jump actions when there is vertically oriented prey straight ahead and both turn actions when there is horizontally oriented prey to a side. SJ works as in Figure 10, assessing whether it is better to step or jump, such that the correct forward action is picked in case TF selects forward movement. LR promotes leftward turns and inhibits right turns when there is horizontally oriented prey to the left and not to the right and vice versa, such that the correct turn direction is picked in case TF selects to make a turn. Blue connections may vary.

By combining recognition of multiple situations into single neurons, the control systems evolved in the main model allow the networks to exploit structural aspects of the environment. Hence we see that when the species is forced to evolve learning ability, these aspects find their way into the implementation of behaviour. It is clear how exploitation of such relations benefit learning. It is also clear that a system without learning ability has little use for such relations. Thus their internalization depends on the need for learning ability and is markedly more prominent in the main model than in the baseline model. This in turn can also explain why functionality becomes highly focused in a small number of neurons: in order to weigh two action choices, the perceptual information pertaining to those choices must be brought together. These observations support the hypothesis put forth in Section 2.

Although it is not in the scope of this research to explore the intricacies of the learning process itself (as mentioned above, the implementation of the learning process is not under the organizing selection hypothesized in Section 2), the connection patterns shown in Figures 10 and 11 suggest the interesting possibility that the nets may be employing perception biases to make useful learning material out of fitness-wise meaningless situations. Assume that there is a 1-cell positioned two steps ahead. During the performance phase a 1-cell in this position could equally well be part of a laterally or longitudinally approached prey, so the situation of a toy object two steps ahead during the learning phase is quite ambiguous. Yet the response of the SJ neuron resembles the response it gives to a laterally approached prey (and comes to trigger the same behavioural response). Similar biases exist for many of the other states encountered during the learning phase. We will not pursue this line of inquiry here, but

it would be interesting to explore if and how such biases relate to the role of imagination in play behaviour. In play behaviour too, fitness-irrelevant situations are often perceived and responded to as fitness-relevant situations, especially where the actual fitness-relevant situations are too scarce or dangerous to drive the learning process.

6. Discussion

The hypothesis presented in Section 2 is closely related to what Godfrey-Smith calls the *environmental complexity thesis*, the idea that environmental complexity is the driving force behind the evolution of cognition [17–19]. Godfrey-Smith identifies Spencer as the first to hold a version of the environmental complexity thesis [18], and we see some of Spencer's continuity between life and mind in Godfrey-Smith's own version of the thesis: "[Cognitive] capacities vary across different types of organism and are not sharply distinguished from other biological capacities, some of which have a "proto-cognitive" character." [19]. However, in his own version of the thesis, Godfrey-Smith steers clear from claims of correspondence between cognitive and environmental organization, opting to defend a weaker version than Spencer. Our results suggest that there is a tenable intermediate position, weaker than Spencer's but stronger than Godfrey-Smith's. This position states that those parts of cognition that evolved under selection pressure for learning ability will tend to express correspondence with the environment.

The organizing effect shown here may also seem related to the Baldwin effect [20], so it seems appropriate to shortly discuss how the effects differ. The interpretation of the Baldwin effect commonly studied in artificial life [21] is that learning ability can accelerate evolution or even allow evolution to find solutions that are extremely hard to find without learning, by "smoothening" the fitness landscape [22, 23]. Learning plays a supportive role, facilitating evolution of traits that by themselves improve fitness. The effect demonstrated in the present paper is different, in that the trait being evolved is learning ability itself. The organization that emerges as a side effect of the evolution of learning is only adaptive in context of the learning ability that it supports. In other words, instead of letting learning ability facilitate evolution of some fit trait, we evolved learning ability to trigger emergence of a trait that facilitates learning. Although closely related (traits supporting learning seem quite susceptible to Baldwinian evolution [9]), these effects do not trivially translate into one another.

Let us also note the described mechanism's relation to the concept of *evolution of evolvability* from the field of bioinformatics. The simplest case of evolution of evolvability can be observed by switching the evolutionary target (the optimal phenotype) back and forth between two options every so many generations. Under these conditions, direct fitness of an individual depends on its phenotype alone, but the fitness of its descendents will additionally depend on how readily the individual's genotype can evolve to a genotype that expresses as the noncurrent target phenotype. Hence lineages that can flick back and forth between the two targets most "efficiently" (e.g., requiring only a small number of common mutations) have an evolutionary advantage. Thus a lineage-level evolution process comes to optimize the genotype's *evolvability* with respect to the environmental dynamic, such that the "regular" evolution process accelerates over time, taking less and less time to move the population from one target to the other. This two-target scenario is explored by Crombach and Hogeweg [24] using gene regulatory networks. The resulting genotypes' propensity to mutate back and forth between the targets with minimal "mutational effort" is a lineage level adaptation to the environment's target switching dynamic and might in a broad sense be said to correspond to it. Kashtan and Alon [25] evolve neural networks and electronic circuits in a more complex task-switching scenario and find modularization and motif formation in the evolved architectures. The mechanism explored in the present paper is essentially the same mechanism, except occurring between a different pair of adaptation processes (evolution and learning instead of lineage evolution and evolution). Depending on the adaptation processes involved, the mechanism may pertain to different ranges of biological phenomena, with cases where learning is involved being particularly pertinent to advanced cognition.

In the field of AI, there is a tendency to view learning and evolution as interchangeable: adaptation is adaptation, be it via advantageous mutations being picked out from the detrimental ones by natural selection or via advantageous behaviours being picked out the from detrimental ones by reinforcement. Evolutionary algorithms are often even classed as a type of learning algorithms. The results presented here suggest that this tendency is detrimental and should be selected against. To understand the evolution of cognition, we should pay close attention to the interactions between learning and evolution, and this requires that we clearly distinguish the various levels of adaptation at work. For a given type of interaction to occur, the nature of the adaptation processes involved may well be immaterial, but even so the same interaction will never occur between a single adaptation process. Furthermore, as stated above, the same type of interaction occurring between different pairs of adaptation processes will pertain to different natural phenomena, making the distinctions between these processes quite important. Let us shortly add another candidate instance of the interaction type we have focused on here.

In this paper we discussed innate correspondence, emerging over the course of the evolution. Advanced cognition makes use of correspondence at a different level as well as hinted at in the introduction: mental representation can be thought of as a form of correspondence that individuals dynamically acquire as they interact with an environment, that is, correspondence acquired via learning instead of evolution. When the adaptation processes involved are evolution and learning, the interaction discussed here does not cover such correspondence. However, the mechanism may be able to account for acquired correspondence if we shift our focus to yet another pair of adaptation processes: learning and second-order learning. We report on our attempts at such extension elsewhere [26, 27].

7. Conclusions

In this paper we discussed a novel theory on the effect of selection for learning on the structure of cognition. We hypothesized that the evolution of learning causes assimilation of environmental structure into neurocognitive structure. Using a simple form of neuromodulation, we built a model in which a form of reinforcement-free learning is made to evolve in neural networks. We found evidence for the hypothesized effect in our evolved networks and considered this evidence in the context of the environmental complexity thesis. We believe that a position in between Spencer's and Godfrey-Smith's is tenable, and that this position provides a useful angle for the computational study of the evolution of neurocognitive organization. We also discussed how the interaction effect demonstrated here can be viewed as an instance of a general pattern of interaction between adaptation processes.

Acknowledgments

The first author thanks Gerard Vreeswijk, Thomas Müller, and Janneke van Lith for many helpful comments.

References

1. H. Spencer, *The Principles of Psychology*, Appleton, New York, NY, USA, 3rd edition, 1885.

2. R. A. Brooks, "Intelligence without representation," *Artificial Intelligence*, vol. 47, no. 1–3, pp. 139–159, 1991. View at Google Scholar · View at Scopus

3. J. A. Fodor and Z. W. Pylyshyn, "Connectionism and cognitive architecture: a critical analysis," *Cognition*, vol. 28, no. 1-2, pp. 3–71, 1988. View at Google Scholar · View at Scopus

4. J. Fodor and B. P. McLaughlin, "Connectionism and the problem of systematicity: why Smolensky's solution doesn't work," *Cognition*, vol. 35, no. 2, pp. 183–204, 1990. View at Google Scholar · View at Scopus

5. B. P. McLaughlin, "Systematicity redux," *Synthese*, vol. 170, no. 2, pp. 251–274, 2009. View at Publisher · View at Google Scholar · View at Scopus

6. R. A. Jacobs, "Computational studies of the development of functionally specialized neural modules," *Trends in Cognitive Sciences*, vol. 3, no. 1, pp. 31–38, 1999. View at Publisher · View at Google Scholar · View at Scopus

7. J. A. Bullinaria, "Understanding the emergence of modularity in neural systems," *Cognitive Science*, vol. 31, no. 4, pp. 673–695, 2007. View at Google Scholar · View at Scopus

8. J. A. Bullinaria, "The importance of neurophysiological constraints for modelling the emergence of modularity," in *Computational Modelling in Behavioural Neuroscience: Closing the Gap Between Neurophysiology and Behaviour*, D. Heinke and E. Mavritsaki, Eds., pp. 187–208, Psychology Press, 2009. View at Google Scholar

9. S. F. Arnold, R. Suzuki, and T. Arita, "Evolving learning ability in cyclically dynamic environments: the structuring force of environmental heterogeneity," in *Proceedings of Artificial Life XII*, pp. 435–436, MIT press, 2010.

10. P. M. Todd and G. F. Miller, "Exploring adaptive agency II: simulating the evolution of associative learning," in *Proceedings of the 1st International Conference on Simulation of Adaptive Behavior*, pp. 306–315, 1991.

11. S. Nolfi, J. L. Elman, and D. Parisi, "Learning and evolution in neural networks," *Adaptive Behavior*, vol. 3, no. 1, pp. 5–28, 1994. View at Publisher · View at Google Scholar

12. S. Nolfi and D. Parisi, "Learning to adapt to changing environments in evolving neural networks," *Adaptive Behavior*, vol. 5, no. 1, pp. 75–98, 1996. View at Google Scholar · View at Scopus

13. E. Robinson and J. A. Bullinaria, "Neuroevolution of auto-teaching architectures," in *Connectionist Models of Behavior and Cognition II*, J. Mayor, N. Ruh, and K. Plunkett, Eds., pp. 361–372, World Scientific, Singapore, 2009. View at Google Scholar

14. A. Soltoggio, P. Dürr, C. Mattiussi, and D. Floreano, "Evolving neuromodulatory topologies for reinforcement learning-like problems," in *Proceedings of the IEEE Congress on Evolutionary Computation (CEC '07)*, pp. 2471–2478, September 2007. View at Publisher · View at Google Scholar · View at Scopus

15. A. Soltoggio, J. A. Bullinaria, C. Mattiussi, P. Dürr, and D. Floreano, "Evolutionary advantages of neuromodulated plasticity in dynamic, reward-based scenarios," in *Proceedings of Artificial Life XI*, pp. 569–576, MIT Press, 2008.

16. L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring expression data identification and analysis of coexpressed genes," *Genome Research*, vol. 9, no. 11, pp. 1106–1115, 1999. View at Publisher · View at Google Scholar · View at Scopus

17. P. Godfrey-Smith, "Spencer and Dewey on Life and Mind," in *Proceedings of the Artificial Life 4*, R. Brooks and P. Maes, Eds., pp. 80–89, MIT Press, 1994.

18. P. Godfrey-Smith, *Complexity and the Function of Mind in Nature*, Cambridge University Press, 1996.

19. P. Godfrey-Smith, "Environmental complexity and the evolution of cognition," in *The Evolution of Intelligence*, R. Sternberg and J. Kaufman, Eds., pp. 233–249, Lawrence Erlbaum, Mahwah, NJ, USA, 2002. View at Google Scholar

20. J. M. Baldwin, "A new factor in evolution," *American Naturalist*, vol. 30, pp. 441–451, 1896. View at Publisher · View at Google Scholar

21. P. Turney, D. Whitley, and R. W. Anderson, "Evolution, learning, and instinct: 100 years of the baldwin effect," *Evolutionary Computation*, vol. 4, no. 3, pp. 4–8, 1996. View at Google Scholar

22. G. E. Hinton and S. J. Nowlan, "How learning can guide evolution," *Complex Systems*, vol. 1, pp. 495–502, 1987. View at Google Scholar

23. R. Suzuki and T. Arita, "Repeated occurrences of the baldwin effect can guide evolution on rugged fitness Landscapes," in *Proceedings of the 1st IEEE Symposium on Artificial Life (IEEE-ALife'07)*, pp. 8–14, April 2007. View at Publisher · View at Google Scholar · View at Scopus

24. A. Crombach and P. Hogeweg, "Evolution of evolvability in gene regulatory networks," *PLoS Computational Biology*, vol. 4, no. 7, article e1000112, 2008. View at Publisher · View at Google Scholar · View at Scopus

25. N. Kashtan and U. Alon, "Spontaneous evolution of modularity and network motifs," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 39, pp. 13773–13778, 2005. View at Publisher · View at Google Scholar · View at Scopus

26. S. F. Arnold, R. Suzuki, and T. Arita, "Modelling mental representation as evolved second order learning," *Proceedings of the 17th International Symposium on Artificial Life and Robotics*, pp. 674–677, 2012. View at Google Scholar

27. S. F. Arnold, R. Suzuki, and T. Arita, "Second order learning and the evolution of mental representation," in *Proceedings of Artificial life XIII*, pp. 301–308, MIT press, 2012. View at Publisher · View at Google Scholar