

The problem with semantic drift:
A close inspection of dedicated learning
mechanisms with evolutionary robotics

Marieke Rohde and Ezequiel Di Paolo
Centre for Computational Neuroscience and Robotics (CCNR)
Department of Informatics, University of Sussex, Brighton, BN1 9QG, UK
Phone: +44 1273 87-2948
Fax: +44 1273 87-7873
{m.rohde,ezequiel}@sussex.ac.uk

August 15, 2006

Abstract

Keywords: Evolutionary Robotics, Value Systems, Adaptation, Learning.
Running Header: The problem with semantic drift

1 Introduction

Any attempt to answer the question ‘What is learning?’ with one sentence is doomed, as it is one of those big terms that cannot be defined in necessary and sufficient conditions, but whose different uses are marked by a ‘family resemblance’, as described by Wittgenstein for the concept ‘game’[23]. One fundamental aspect in the concept of learning that spans most uses of the term is, however, that it encompasses a change in behaviour, usually a change to the better, and usually a change that is comparatively long lasting. In dynamical systems terms, it appears intuitive to think of learning as a dynamical process on a time scale slower than the dynamics of sensorimotor behaviour.

A very common approach towards explaining learning is to propose two different mechanisms for these two behavioural time-scales, one mechanism for fast sensorimotor behaviour, and another one, separate and slower, for the modulation of this behaviour in a top-down way. In traditional artificial intelligence, this distinction between the “learning element, which is responsible for making improvements, and the performance element, which is responsible for selecting external actions” ([16], p. 525) is very explicitly made, it underlies the sub-field of machine learning. In this view, what is learned is separate from what learns, and, therefore, what learns improves behaviour, but is in itself, not subject to improvement. But also in less conservative synthetic approaches, such as neural networks or evolutionary approaches, such a distinction is widely adopted. For instance, in neural networks, learning is typically realised as a slow convergence of connection weights according to a separate and fixed learning rule, as, e.g., in backpropagation learning[15]. Similarly, evolutionary approaches typically investigate the benefits of hardwiring a behaviour-modulator that realises sensitivity to lifetime experience, as opposed to hard-wiring behaviour directly (e.g., [1, 11, 20]).

Indeed, even though the outlined functional and structural separation of learning and behaviour generation is by no means *a priori* necessary, it seems next to impossible to find synthetic approaches to learning that do not implement this divide. The only exceptions to the rule that we can think of are recent evolutionary robotics studies that provide the existence proofs for alternatives to the presupposition of separate mechanisms ([24, 21, 9, 19]). We have the impression that it is not widely recognised that such a separation is not strictly logically necessary, and, furthermore, that its presupposition entails further commitments.

In this paper, we want to argue and illustrate, with deliberately simple evolutionary robotics simulations, that the assumption of a separation between learning and behaviour generation has certain unpleasant consequences, if looked at in the context of dynamical and embodied interaction with an environment. We believe that it is helpful for researchers to be aware of these consequences, so they can address them. In particular, we argue that any involvement of sensorimotor behaviour in the mechanisms of learning will, as a consequence to the circular causality thereby created, lead to *semantic drift* and that the question of how learning mechanisms are protected against semantic drift requires to be

answered. As example learning theory we discuss ‘value system architectures’ as proposed by Edelman et al. in the theory of neuronal group selection (TNGS, e.g., [5]). Laudably, in TNGS, other than in most learning theories, the mentioned presupposition of a divide is made explicit within the context of situated and embodied adaptation of behaviour. For this reason, this theory is popular in the robotics community, where it is, e.g., advocated by Pfeifer and Scheier [14], who argue that self-supervision through value systems is essential to direct processes of self-organisation in autonomous agents. And for the same reason, it is very suitable for our critical investigation of the entailments of the divide between what learns and what is learned in an embodied system.

For we have been misunderstood before, when arguing our points to peers, every reader that feels either bored or offended after reading our argument in section 2 is kindly asked to not put our paper aside before having read section 3, in which we try to preempt misunderstandings. Of course, the results from research fields that presuppose the described divide are invaluable, and it is certainly not the objective of this paper to downplay their importance or benefit for the field of AI and cognitive science. Having recognised the merits of adopting this simplifying assumption, we do believe, however, that the points we are making are of crucial importance and have failed to be answered. Any reader who is intrigued and free of reservations against our argument after reading section 2 is encouraged to move on to the following section 4 immediately.

In section 4, we present a series of minimalist evolutionary robotics simulations to illustrate our arguments. The simulation implements a caricature version of value system architectures as proposed in TNGS, in which agents are evolved to seek light and, at the same time, to generate an estimate of their own performance (i.e., fitness). Since our work here presented is, in the first place, conceptual work, we keep the complexity of the simulation models as low as possible, so the results to our study are very easy to understand. It is not our objective to model new behaviours or generate new problems, these experiments are evoked to demonstrate how architectures that introduce a divide between learning mechanism and what is learned are not guaranteed to work without adopting additional assumptions. In a first experiment, we investigate the relation between sensorimotor behaviour and the generation of a value signal. We then go on to study the suitability of such a signal as internally generated reinforcement signal.

The first set of experiments has a predominantly negative interpretation, in that it criticises without proposing alternatives. The second set of experiments presented in section 5, in contrary, aims to point out how learning in the light of certain empirical evidence can be investigated whilst minimising prior assumptions about the relation between structure and function. Agents are evolved to solve a reinforcement learning task, and, at the same time, to incorporate a value signal into the network dynamics. In this experimental set-up, the functional role of the value signal for the reinforcement learning is underspecified and can be investigated and explained *post factum*. [ONE SENTENCE ABOUT THE EXPERIMENTAL RESULTS, IF THEY ARE THERE BEFORE THE END OF THE CENTURY] [ANOTHER SENTENCE ABOUT HOW THIS IS

The concluding section 7 brings together the conceptual argument with the results from both simulation studies. It evaluates the presented findings in the larger context of the study of learning and adaptation, and how they relate to existing synthetic and empirical work. It will also draw conclusion for the issues of neural correlates and hybrid models in general, and propose ways in which the presented research can be improved and extended.

2 Embodied behaviour, disembodied learning

Many models of adaptivity and learning in artificial agents aim to accommodate demands expressed for embodiment and situatedness of behaviour. An agent that can adapt its behaviour on the basis of the experiences it makes in a variable environment is more flexible than a hard-coded agent. It is not necessary for the designer of a learning agent to predict each and every possible situation and the appropriate behaviour to go with it. Instead, usually, abstract criteria and modification rules are included into the agent architecture. They extract cues from the variable environment about what is good or what is bad (be it explicit teacher feedback, as in supervised learning, bipolar feedback signals, as in reinforcement learning, or structural properties of the data, as in unsupervised learning) and adjust the agent’s part in the action–perception–loop accordingly. Our model learning theory, TNGS, has a particular instantiation of this principle of cue extraction: These models feature a *value system*, which generates a bipolar performance signal in response to behavioural success, which is then used as selection criterion for a Darwinian–style evolutionary process of natural selection, but within an agent’s lifetime, that reinforces successful behaviour by strengthening the participating synaptic connections (value guided learning). For instance, a value system for reaching would become active (“good”) if the hand comes close to the target [18]. This activity of value systems can be seen as the internal generation of a reinforcement signal.

As pointed out earlier, learning mechanisms in themselves are typically assumed not to adapt to changes, the principles they implement are — implicitly or explicitly — thought to be generally applicable and not sensitive to changes to the body or the environment. In this sense, even though they help dealing with variability in the body or the environment, the learning mechanism itself can be seen as disembodied, as outside and beyond affection through aspects of embodiment and situatedness. This presupposed rigidity is explicitly endorsed in TNGS: Value systems are thought to be “already specified during embryogenesis as the result of evolutionary selection upon the phenotype” [18] and judge according to prespecified rules (“simple criteria of saliency and adaptiveness” [18]). Hence, the functionality of mechanisms of “adaptability and flexibility in the presence of changing biomechanical properties” ([18]) are assumed to be phylogenetically constant. They are assumed to be, in themselves, insensitive to changing biomechanical properties.

Some authors hold it possible “that different value systems interact, or that

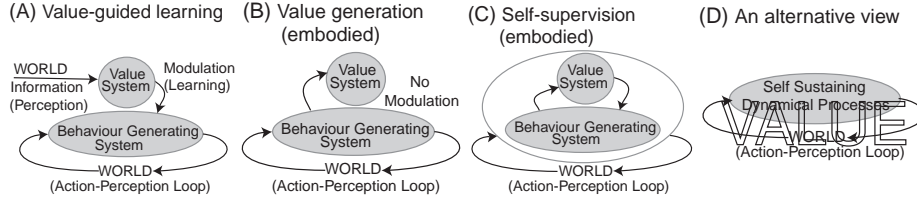


Figure 1: Schematic view of value system architectures as proposed (A), of the variations tested in our first (B) and second (C) set of experiments and an alternative view on values (D).

hierarchies of specificity might exist” [18], i.e., they do leave some room for variability in value system functionality. However, what is clear is that no value system can judge on *its own* performance because this would logically undermine the separation of learning mechanism from what is learned. Thus, if we do not want to enter a *regressus ad infinitum*, we have to acknowledge that, at some point, the hierarchy of value systems must end, unless we want to endorse a mystical supplementary force that adapts value systems and maintains their functionality. An explanation of behavioural plasticity will be incomplete, if it proposes a plastic learning mechanism but does not explain its plasticity, which is practically just pushing unexplained plasticity one level up.

Another consequence of introducing the divide is that the behaviour generating mechanisms themselves are agnostic towards the cues that the learning system extracts. They just do whatever they are doing, the learning system manages their variability and decides whether it is any good. The behaviour generating systems blindly obey the learning system, they do not have a say. A schematic diagram of this kind of architectures can be seen in figure 1 (A): We see an agent, embodied, embedded and behaving within a closed sensorimotor loop, and a learning module which monitors and adjusts this embodied and embedded activity, separately, in a top-down manner, and which is, in itself, disembodied, in the sense that its functionality is not adaptive and not sensitive to bodily or environmental variables.

Such circuits can work in particular settings, as demonstrated, e.g. for the case of value system architectures in an example robotic application by Verschure et al. [22]. However, we believe that there are limits to the suitability for such approaches to explain adaptivity, and we are not the first ones to have recognised these limits. The following discussion adopts arguments by Julie Rutkowska [17], who argues that “[increased] flexibility requires some more general purpose style of value” [17], and Susan Oyama [13], who has criticised the nature/nurture divide in biology and psychology in a way that we conceive as analogous to the divide between learning and the learned in artificial agent research.

The disembodied nature of value system architectures leads to Rutkowska’s question whether they constitute a “vestigial ghost in the machine” [17]. She laments that, due to the built-in evaluation criteria, their semantics are restric-

tive. A similar limitation is pointed out by Pfeifer and Scheier, who describe a “trade-off between specificity and generality of value systems” [14]: A very specific value system will not lead to a high degree of flexibility in behaviour, while a very general value system will not constrain the behavioural possibilities of the agent sufficiently. By virtue of their disembodiment, such systems are subject to the criticisms uttered about traditional disembodied artificial intelligence architectures (e.g., [4, 12, 14]). They are rigid and non-adaptive, their functionality relies on the intact functionality of dedicated input and output channels, which makes them vulnerable, and, even if the systems they control are more flexible, the learning systems themselves can only deal with scenarios that could be foreseen when they were designed.

Sceptical voices will maintain that, despite these points, empirical evidence favours theories of learning that include a functional and structural separation of learning mechanism and what is learned. TNGS, for example, is backed with empirical evidence about a correspondence between salient events in the environment and the activity of cell assemblies in the brain stem and the limbic system that modulate synaptic changes in the cortex[7]. Similarly, the actor critic model, a temporal difference reinforcement learning method has been said to be biologically plausible and possibly implemented in the basal ganglia [2]. This list could be continued, and such empirical findings clearly point towards a major involvement of the described neural assemblies in learning.

We do not want to deny that there is a major involvement of certain brain areas and certain physiological mechanisms that change on a slow time scale in the realisation of behavioural learning. What we want to stress is that a major involvement, a strong correlation, is not an explanation in itself, in the same sense that registering the strong involvement of the gas pedal in a car’s acceleration behaviour is not an explanation of how a car drives. To know about the gas pedal may help to learn, in an instrumental way, how you can make a car go fast, but, if it does not involve the motor, combustion, the crankshaft, it is not an explanation. If the car breaks down, knowing about the gas pedal will only help in a very limited number of cases. Furthermore, these models, valuable as they may be in understanding particular instances of behavioural adaptivity in situations that rely on phylogenetic constancies, are confronted with serious problems as soon as more complex, temporally extended, abstract or high level adaptation processes are addressed. [THESE PROBLEMS ARE WHAT IS KNOWN AS CREDIT ASSIGNMENT PROBLEMS ISNT IT?]

Additionally, a living organism is, other than a car, in constant material flux. Can we treat parts of an organism as *a priori* functionally and structurally specified and exempted from ontogeny? Oyama [13] points out that in biology “[developmental] information itself [...] has a developmental history” and is “developmentally contingent in ways that are orderly but not preordained” [13]. Insofar, it appears misconceived to speak of traits as inherited, as opposed to acquired. She goes as far as stating that “[n]or can phenotypic features be divided into those that are programmed or biological and those that are not” [13], an argument that clearly and directly applies to the described case of learning mechanisms that are supposed to be evolutionarily hard-coded.

The proponents of separate learning mechanisms will admit that, maybe, there is some variability, some ontogenetic change and bodily influence. But does that imply that the story could not *roughly* go the way proposed? We believe that any isolated symbolic structure, any vestigial ghost in the machine, will suffer the same problems in a variable bodily environment that a full blown ghost in the machine suffers in a variable external environment. ‘Symbol grounding’[8] supposedly takes place, if a symbol processor is hooked up to a real world context through sensorimotor couplings. But will slight alteration of this context not result in an alteration of the semantic grounding? It can, as we show in our simulation experiments (section 4). Such constant slight meaning changes, resulting from a variable physiological context, is what we term *semantic drift*. We believe that semantic drift is a problem for hybrid approaches that introduce a fixed and *a priori* correspondence between a certain function and a certain mechanical structure, which is embedded in a variable context, but is not variable in itself.

How is semantic drift countered? We do not say that this problem cannot be addressed from within a proposed learning theories – all we want to say is that most approaches have, so far, failed to do so. It is easy, but slightly unsatisfactory to functionally divide a task and assign sub-functions one-on-one to mechanical structures, simply assuming that these structures will autonomously make sure they fulfill their duty well. An explanation that addresses the processes that sustain functionality within the global context, inspite of variability of the external and internal environment, will be so much richer.

The simulation experiments we present in this paper investigate these issues in minimal controlled settings. We want to emphasise again that these experiments are simply illustrating the entailments of assuming a divide between learning mechanism and what is learned argued verbally above. They will not lead to surprising new results or theories of learning. A mobile agent is designed through artificial evolution to perform simple phototaxis and, at the same time, to generate a signal that corresponds to its level of performance. In a first set of experiments, a value signal is generated that has no effect on the network dynamics, there is no *a priori* need or function associated with this estimate, it simply serves us as analogy with the aforementioned evidence about brain structures whose activity corresponds to salient events in the environment. With this experiment, we question whether you can conclude from a correlation between a meaningful variable and activity in a separate neural structure that this structure implements a certain function (as in Fig. 1 (A)), or if the sensorimotor context can, nonetheless, contribute to the function (Fig. 1 (B)). In a second set of experiments, the internally generated value signal is fed back into the neural dynamics of the agent (Fig. 1 (C)). With this experiments we address the problem of semantic drift. We emphasise the consequences of the reciprocal causal links that go in both directions, not only top-down from the value system to the behaviour generating network. In the section 5 we present an alternative methodological approach to problems of reinforcement learning that does not try to localise value in part of the agent architecture *a priori* (Fig. 1 (D)).

3 Don't get me wrong

It is impossible to preempt every misunderstanding that could possibly arise, however, we still want to try our best to avoid losing our readers along the way, or any unnecessary ill-feeling. We first want to make clear that we are not antagonising researchers from other schools or disciplines. Then, we want to make clear that our points are, nonetheless, of wide applicability and crucial interest to the learning community.

- *Neuroscience.* Our points about the problems with reducing function to neural structure do not have to be interpreted as a criticism of neuroscience as a research agenda, which would be very counterproductive and stupid. We think that neuroscience, as an instrumental science, is invaluable for both therapeutic practice and cognitive science. We have a huge respect for people who take up the challenge to explain the most complex and least understood organ of human anatomy, be it with hands-on empirical research or with computational modelling as a part of theory building. Given the complexity of the brain and the technological possibilities today, it is absolutely clear that only very particular — and this frequently means local — aspects of the brain and its dynamics can be investigated at any point in time, and that simplifying assumptions have to be adopted. However, it is important to be clear about the assumptions one adopts and what they mean in the context of global brain-body-environment interaction, as many neuroscientists do (e.g., [?, ?, ?, ?] Varela engel koenig singer).
- *Learning.* In a similar way, our discussion is not in any way opposed to research in learning that presupposes a divide between learning mechanism and what is learned. It is just important to be aware that this separation is not strictly necessary, and to interpret the findings accordingly. In some situations, such a model may be seen as description rather than as explanation, in others, it may be convenient to postulate some further processes or mechanisms yet to be explained. In a general theory of learning, however, assumptions about the how learning behaviour and mechanisms of learning relate should not be primitives, built into the explanation of learning behaviour. They should result from the investigation of the learning agent.
- *Correlations.* The evidence for correlation between measurable variables in the brain and meaningful variables of human-environment interaction (such as the correlation between saliency of events and activity in cell assemblies in the brain stem and the limbic system [7]) is huge, and hugely interesting, if looked at in the right light. The problems arise only, if such correlations are interpreted, consciously or unconsciously, as *internal representations* in the classical computationalist sense. This means that they would play the role of a signal, and just a signal, to stand in for some meaningful variable to a higher-brain-area-homunculus. To recur

the example of the car: To think of the state of the gas pedal as a velocity signal on the basis of correlation alone is neither justified nor useful for explanation .

- *Empirical truth or theoretical possibility?* From the simple experiments presented here, nothing can be inferred on how learning works in humans or animals. Explaining how organisms or their brains work is not our ambition. What we want to do is to point out the consequences of an important theoretical possibility — namely that the mechanisms of behaviour generation and the mechanisms of learning may not be structurally or functionally fully separate. Whether or not this theoretical possibility is true for one biological species or the other remains to be shown through empirical research. Theories that rely on the presupposition of this divide, however, will not serve to argue in favour of the existence of the divide, because the alternative has never really been considered.

The agreement of our critical points with most scientific practice in general, however, does not mean that it is without consequence for most of the scientific practice.

- *Not just value system architectures.* In our argument, we have done our best to make clear that we did not chose TNGS as a particularly objectionable theory, but as a very suitable and well spelled-out token theory in the class of learning theories that assume a functional and structural divide between learning mechanism and behaviour generating mechanisms, a class in which most synthetic models of learning fall. Our criticism, therefore, is not tied to TNGS with its particular merits and demerits.
- *More than just noise.* If we point out the difficulties with a clear cut separation between behaviour generation and adaptation, an expectable response is that even though things may be not that black and white in biology, these theories still roughly capture the way things work, because the impact of structural changes and complex feedback loops is practically noise that can be filtered out. We do not believe that the problem of semantic drift is so negligible, because it applies to the level of behavioural function. Whatever filter would be applied in these architectures, it would have to know what is signal and what is noise, which requires some rather elaborate *cognitive* capacity from this filter, which is hard for us to imagine.
- *Learning to learn.* A very similar point can be made about proclaiming adaptability of learning mechanisms, but not specifying the details of how they adapt. If the need for adaptive plasticity of learning mechanisms is acknowledged, than this can only happen through other learning mechanisms, if we want to keep up the divide between learning mechanism and what is learned. Such a mechanism would, again, have to be a very powerful mechanism to be able to tell that something is wrong in the

functionality of the learning mechanism and to be able to do something about it. Simply stating that there is ways of maintaining the meaningful functionality of such circuits without stating the mechanisms of their maintenance is not an explanation, it pushes the problem of adaptivity one level up. Also: If such powerful mechanisms are possible, why do they not take care of behavioural adaptation in the first place?

- *An obvious point?* For everyone who sees our argument, the results of the simulation experiments we present will not be surprising. The model is deliberately simple, because it is meant to be understood and to illustrate the obvious. It demonstrates the consequences that follow from presuming a structural divide between learning mechanism and behaviour generating mechanism without making further reference to the relation between learning mechanism and embodied interaction with the environment. If these consequences are so obvious and trivial, the question that arises is: Why have these pressing issues not been more frequently addressed?
- *Low level and high level.* It is very tempting to conclude from the fact that the simulation experiments we present are very simple and low level that our arguments would be restricted to low level sensorimotor behaviour. This conclusion, however, is not justified. For a start, empirical evidence shows the direct influence that sensorimotor properties have on even the most high level cognitive capacities like human conscious experience (see work on sensorimotor contingencies O'Regan) or [some faculty mathematical problem solving and lakoff nunhez?]. Furthermore, as we pointed out earlier, it seems that *a priori* specified learning circuits, with their restrictive semantics that require situations to be phylogenetically predictable, seem much more likely to suffer from a restriction to very simple situations of behavioural adaptation. An approach that does not aim to demarkate a behaviour space that is searched according to prespecified rules of adaptivity is much more likely to account for genuine novelty, sense-making and open ended development.

4 Correlation and Meaning

4.1 The Model

The model is deliberately minimalist. It does not aim to model actual brain structures, as the cited models, it serves to illustrate a conceptual argument.

A circular two-wheeled agent of 4 units diameter is designed by evolutionary search to perform phototaxis. The control networks evolved are continuous time recurrent neural networks (CTRNNs, e.g. [3]) with variable size and structure (see below). The dynamics of neurons n_i in a CTRNN of N neurons are governed by

$$\tau_i \frac{da_i(t)}{dt} = -a_i(t) + \sum_{j=0}^N c_{ij} w_{ij} \sigma(a_j(t) + b_j) + I_i \quad (1)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the standard sigmoidal function and I_i is the external input to n_i . The weights $w_{ij} \in [-8, 8]$ from n_j to n_i , the bias $b_i \in [-3, 3]$ and the time constant $\tau_i \in [16, 516]$ are determined by a genetic algorithm (GA). C is the $n \times n$ connectivity matrix with $c_{ij} = 1$ if there is a connection from n_j to n_i and $c_{ij} = 0$ otherwise.

The agent has two sensors $S_{L,R}$ with an angle of acceptance of 180° , which are oriented towards $+60^\circ$ and -60° , with added uniform directional noise $\in [-2.5^\circ, 2.5^\circ]$. Their activation is fed into input neurons by $I_{S_i}(t) = Sg \cdot S_{L,R}(t)$ with Sg evolved $\in [0.1, 50]$ and $S_{L,R}(t) = 1$ if the light is within the sensory range of $S_{L,R}$ at time t and $S_{L,R}(t) = 0$ otherwise. Note that the binary character of the light activation makes the estimation of the distance to the light non-trivial. The motor velocities are set instantaneously at any time t by $M_{L,R}(t) = M_G \cdot (\sigma_{M_{i+}}(t) - \sigma_{M_{i-}}(t)) + \varepsilon$ where M_G is the motor gain $\in [0.1, 50]$. $\sigma_{M_{i\pm}}(t)$ is the neural output of one of the two neurons controlling $M_{L,R}$ and $\varepsilon \in [0, 0.2]$ is uniform noise. A fifth output neuron generates the performance estimate $E(t) = \sigma_{M_5}(t)$.

The connectivity C and the size of the network is partially evolved. Connections to input neurons or from output neurons are not permitted. Input neurons can project to output neurons and to hidden neurons, hidden neurons can project to other hidden neurons and to output neurons. The network can have varying numbers (0–5) of hidden neurons. In experiments where the value signal E is integrated into the network dynamics (Sect. 4.4), the estimator neuron changes status to become another interneuron. In some experiments, parts of the network structure and parameters were excluded from continued evolution at a certain stage.

Parameters for the control network are evolved in a population of 30 individuals with a generational genetic algorithm with real-valued genes $\in [0, 1]$, truncation selection ($\frac{1}{3}$), vector mutation [3] of magnitude $r = 0.7$ and reflection at the gene boundaries. The sensor gain S_G , the motor gain M_G and the time constants τ_i are mapped exponentially to the target range. The existence or non-existence of hidden neurons and neuronal connections is determined by the step functions $x > 0.7$ and $x > 0.6$ respectively. All other values are mapped linearly to their target range.

In every evaluation, the robot is presented with a sequence of 4–6 light sources that are placed at a random angle and distance $\in [40; 120]$ from the robot. Evaluation trials last $T \in [3000, 4000]$ time steps. They are preceded by $T' \in [20, 120]$ simulation time steps without light or fitness evaluation, to prevent that the initial building up of activity in the estimator neuron follows a standardised performance curve. Each light is presented for $t_i \in [\frac{T}{5} - 100, \frac{T}{5} + 500]$ time steps. The network and the environment are simulated using the forward Euler method with a time-step of 1 time unit.

The fitness function

$$F(i) = F_D(i) \cdot F_E(i) + \varepsilon F_D(i) \quad (2)$$

is basically the product of a term assessing the behavioural performance, i.e., the distance to the light ($F_D(i)$) and another one assessing the estimation of this be-

havioural performance by the agent ($F_E(i)$). The second summand ($\varepsilon = 0.001$) is simply included to bootstrap the evolution of behaviour, as the coevolution of light seeking and estimation of performance from scratch is difficult for evolutionary search. The first term ($F_D(i)$) is given by

$$F_D(i) = \frac{1 - M^2}{T} \int_0^T \max\left(0, 1 - \frac{d(t)}{d(t_0)}\right) dt \quad (3)$$

with $M = \frac{0.125}{T} \int_0^T \frac{M_L(t) - M_R(t)}{M_G}$. $d(t)$ is the distance between robot and light at time t and t_0 the last displacement of the light source. This term is fairly straightforward and has been adopted from [?]. It rewards fast approach behaviour and punishes the robot for turning.

The fitness estimate F_E , however, has gone through a long but necessary process of refinement and complication. It is given by

$$F_E(i) = \sqrt{\max(0, F_{E1}(i)) \cdot \max(0, F_{E2}(i))} \quad (4)$$

The two factors F_{E1}, F_{E2} denote estimates of the performance and its change over time respectively.

$$F_{E1}(i) = \frac{e(\bar{d}, d) - e(E, d)}{e(\bar{d}, d)} \quad F_{E2}(i) = \frac{e(0, \dot{d}) - e(\dot{E}, \dot{d})}{e(0, \dot{d})} \quad (5)$$

with $e(x, y)$ the sum of squared error $e(x, y) = \int_0^T (x(t) - y(t))^2 dt$. \bar{d} is the average of $d(t)$ during each trial. $\dot{d}(t)$ and $\dot{E}(t)$ are the derivatives of $d(t)$ and $E(t)$ averaged over a sliding time window $w = 250$ time steps (interval borders for $e(x, y)$ have to be adjusted accordingly).

Why this complicated term instead of a simply using the sum of squared error of the generated performance estimate signal ($e(d, E)$)? The problem is one of local maxima in the fitness landscape. To generate a constant signal that corresponds to the average performance is a trivial, but rather high scoring strategy, even if more variability is introduced in the task. Therefore, the fitness function had to be altered such that the sum of squared error had to be *better* than the best constant output (i.e., \bar{d}). This led to the term F_{E1} . However, this repair only guided evolution into the next local maximum: A standardised performance curve that would start at 0 and lead up to the average performance \bar{d} . Therefore, the second term F_{E2} was introduced which requires the estimate to follow the changes in distance ($\dot{d}(t)$) with a higher accuracy than a constant derivative of 0. This alteration then resulted in the evolution of interesting, non-trivial estimation behaviour.

The evaluation of a network i on $n = 6$ trials is given by

$$F(i) = \sum_{j=1}^n F_j(i) \cdot 2^{-(j-1)} \cdot \frac{1}{\sum_{j=1}^n 2^{-(j-1)}} \quad (6)$$

where $F_j(i)$ gives the fitness on the j^{th} worst evaluation trial for individual i , which gives more weight to worse trials and thereby rewards the generalisation capacity of the evolved networks.

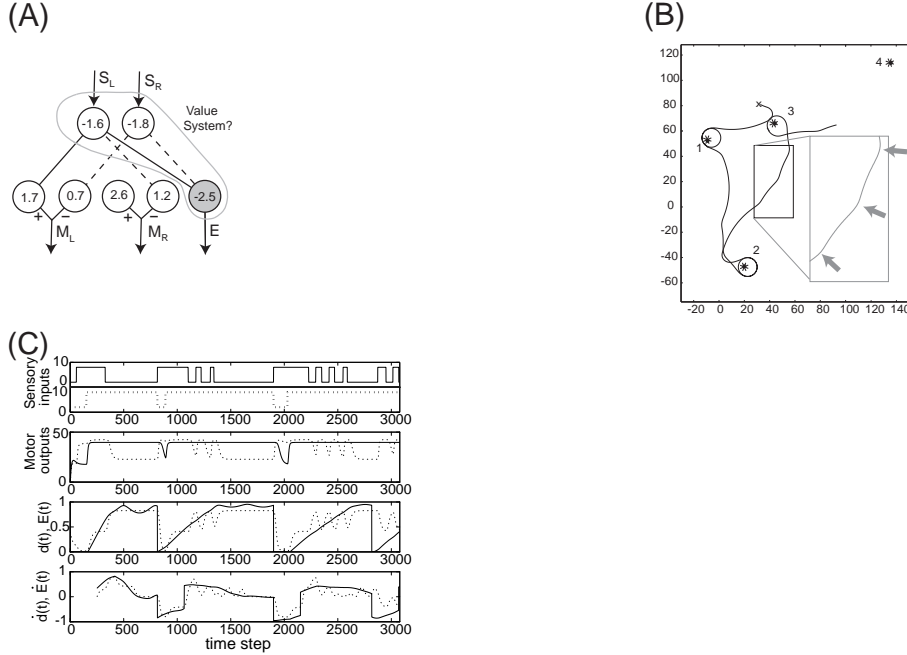


Figure 2: (A) The distance estimator network (θ in neurons, dotted lines inhibition, solid lines excitation). (B) Trajectory following four presentations of light sources. Arrows indicate the punctuated turns during $t = 2200 - 2700$ (see text). (C) The evolution of different variables over time in the same trial (Top to bottom: $S_{L,R}$, $M_{L,R}$, $d(t)$ vs. $E(t)$, $\dot{d}(t)$ vs. $\dot{E}(t)$).

4.2 Results

4.3 Generating a Value Signal

In this section, we describe and analyse an individual evolved agent. It was selected because of its simplicity and because its way of estimating performance is representative for the most frequently evolved strategy.

The network evolved (Fig 2, (A)) does not have hidden neurons, recurrent connections or slow time constants, i.e. its behaviour hardly relies on internal state and its complexity is minimal, even within the already restricted range of possibilities. For rhetorical reasons, we start with the description of the value system, before we describe the light seeking behaviour.

The neural structures participating in the generation of the value signal are just the two input neurons and the estimator neuron, so if anything, we would have to call this sub-system the value system. In the absence of light, or if the network receives input only on its right light sensor ($S_R = 1, S_L = 0$), it estimates $E \approx 0$. If light is perceived with both sensors, it estimates $E \approx 0.5$, and if the network receives input only in its left light sensor ($S_R = 0, S_L = 1$),

the estimate reaches its maximum of $E \approx 0.8$. The judgment criteria of this value system can thus be described as “seeing on the left eye is good, seeing on the right eye or not at all is bad”. Intuitively, these rules do not make sense. Nevertheless, as we can see in Fig. 2 (C) (bottom two plots), both $E(t)$ and $\dot{E}(t)$ (dotted lines) follow with amazing accuracy the actual values $d(t)$ and $\dot{d}(t)$ (solid lines), particularly if we remember the poor sensory endowment of the agent.

The agent’s light seeking behaviour is realised by the network minus the estimator neuron. In the absence of sensory stimulation, the agent slowly drives forward, slightly turning to the right. If $S_R = 1$ and $S_L = 0$, the “brake” on the left motor M_L is released, which leads to a sharper turn to the right. If $S_R = 0$ and $S_L = 1$, the “brake” on the right motor M_R is released, which makes the agent turn to the left. If light is perceived with both sensors, the agent releases both “brakes” and drives almost straight, slightly drifting to the right. In combination (Fig. 2 (B)), upon a presentation of light, these four behavioural modes lead to the following sequence of actions: 1.) A scanning turn to the right, until $S_L = 1$. 2.) A quick approach of the light from the right side. 3.) counter clockwise rotation around the light source. While the agent approaches the light source, it keeps bringing the light source in and out the sensory range of S_R (compare the rhythmically occurring drops of sensory and motor activity in Fig. 2 (C)). This strategy results in the chaining of nearly straight path segments in the approach trajectory, separated by punctual left turns (arrows in Fig. 2 (B)).

We now return to the agent’s value system. The estimator neuron M_5 outputs $E \approx 0$ if $S_L = 0$. The reason for this is that during the entire approach behaviour $S_L = 1$, and therefore $S_L = 0$ implies that the light has not yet been located, which only happens in the beginning of the trials if the agent is far away from the light source. During the nearly straight path segments, $S_L = S_R = 1$, which leads to $E \approx 0.5$, i.e. an intermediate estimate for an intermediate approach stage. While the agent cycles around the light source, $S_R = 0$ and $S_L = 1$, and the value system produces its maximum estimate, expressing that the light source has been reached. Notice also that the straight path segments which correspond to $E \approx 0.5$ become shorter as the agent comes closer to the light. Therefore, even though the value system has just three modes of output, its evolution over time can express a more gradual change in distance, if averaged over a time window: The average output increases with decreasing distance to the light.

Another event worth discussing in the trial depicted in Fig. 2 (B) and (C) occurs after the last displacement of the light source ($t > 2800$): As the displacement happens to bring the light source in the left visual field of the agent, it immediately enters the oscillating approach mode and its estimate therefore poorly corresponds to the actual distance measure which drops to 0. This dissonance can be seen as inevitable error due to the limited possibilities of the agent. However, we prefer to see it as superiority of the evolved estimator over the distance measure as a measure of performance: The comparably high output expresses the agent’s justified optimism to be at the light source soon.

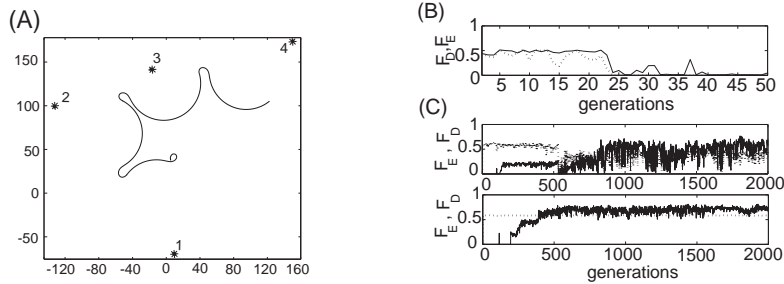


Figure 3: (A) Light-avoiding trajectory of an agent after 50 generations of value guided learning. (B) The degeneration of light seeking performance F_D (solid line) and estimation performance F_E (dotted line) over time (same experiment) (C) Examples: F_E (solid) and F_D (dotted) in coevolving phototaxis (top) and fixed phototaxis (bottom)

Such discrepancies between meaningful judgment signals generated by the agent and *a priori* specified performance measures were one of the key difficulties in designing the experiments. Even with the highly refined and complex fitness measure F_E (4), sometimes, “good” solutions in terms of the experimenter’s perception were replaced with less sophisticated ones by automated selection.

4.4 Value Guided Learning

Value systems are the proposed neural structures to guide ontogenetic adaptation. Can such mechanisms work if the value system is properly embodied? To investigate this question, we conducted another simple simulation experiment, in which the evolution of the robot controller is seen as the analogue of ontogenetic neural Darwinism as proposed in TNGS. The only parameters that evolve in this experiment are the strengths of the three synaptic connections from sensors to motors in the agent presented in the previous section (compare Fig. 2 (A)). The fitness measure F is substituted for the performance estimate $E(t)$. It is important to notice that in this set-up, the value system does not evolve, it just guides the evolutionary change of the synaptic weights to reinforce whatever behaviour leads to a high performance estimate $E(t)$.

Figure 3 (B) illustrates how with an embodied value system, value guided learning quickly results in a deterioration of light seeking behaviour, even though synaptic weights are just minimally altered. What the “value system” rewards is simply activation of the left light sensor but not the right. That this judgment means good light seeking behaviour during embodied interaction is a contribution of the sensorimotor context, and this meaning is removed if the system is functionally separated from the sensorimotor context. The gradual change of behaviour results in what we call “semantic drift” of the value system, i.e. the behaviour it rates as successful quickly ceases to be phototaxis (Fig. 3 (A)).

We see that the functional integration of the value system into the sensorimotor loop has far reaching consequences for the role this value system can play in the adaptation of behaviour dynamics. The reciprocal causal connections between behaviour generating system and value system undermine the idea of the value system as a top-down modulator of behaviour. But if the function of a neural structure whose activity we, as observers, can interpret as performance signal is not actually a value judgment, what could it be? This question is an open issue. One answer has already been given in Sect. 4.3 of this paper: Such a correspondence could be purely epiphenomenal and not bear any functional role in the generation of behaviour.

In an initial attempt to further investigate this question, we evolved agents in which the estimator neuron has the status of an interneuron and can project to other neurons. The most common structure we find in these networks is an excitatory self-connection in the estimator neuron that improves the estimation performance, but not phototaxis. In some of the networks that realise the same strategy described in Sect. 4.3, light seeking crucially depends on the activity of the estimator neuron. It serves to inhibit the right motor, as its activity is roughly in inverse correlation with the activity of the right sensor, and thereby takes part in inducing left turns if the light goes out of the right visual field. Its function is simply to relay and invert the right sensory signal. There is no end to the possible functions a “value system” could serve in the control of an embodied and situated agent. What the presented findings show is that the correspondence of neural activity to a behaviourally meaningful variable may well be plainly accidental.

4.5 The Evolution of Value Systems

Comparing the agents evolved to estimate value and seek lights to agents evolved to achieve just phototaxis (i.e. $F(i) = F_D(i)$), it turns out that the light seeking behaviour in agents that are evolved to estimate their performance is clearly suboptimal. Our first hypothesis to explain this phenomenon was a trade-off between the ability to perform judgments and the ability to find light quickly.

To test this hypothesis, we seeded evolution with successful light seeking agents and evolved combined light seeking and judgment behaviour on top, comparing conditions in which the sensorimotor behaviour was either fixed or continued to evolve with the value system. We expected the latter to be fitter, because the light seeking behaviour could be changed by evolutionary search to allow better estimation of performance. To our surprise, we found that both F_D and F_E were on average higher in the agents with fixed sensorimotor behaviour¹. If good light seeking and good value estimation are possible at a time, why does the evolutionary search not find this solution? If we have a closer look at how the F_E and F_D component evolve in example evolutionary runs (Fig. 3, (C)), we see that the coevolutionary scenario (top) is much more noisy and good solutions

¹However, one of the seeded phototactic agents applies a strategy for phototaxis that does not seem to allow the estimation of performance. This suggests that there is at least some need for sensorimotor behaviour to accommodate judgment.

repeatedly deteriorate. Apparently, in the presented set-up, a good estimation of the agent’s performance is very sensitive to behavioural noise and can only exceed a certain level if the sensorimotor coupling is completely fixed. This explains why value guided learning leads to such a rapid and devastating decay of behaviour: The noise sensitivity of value estimation accelerates semantic drift.

5 The second set of simulations

6 Discussion

Summarising the results from our simulation experiments, we presented an agent in which the capacity to judge on its level of performance with respect to a certain task crucially relies on the sensorimotor behaviour through which this task is realised. Without this sensorimotor context, the neural structure producing the performance estimate is meaningless, and if sensorimotor behaviour does not accommodate the need to estimate the level of performance, such judgment is only possible to a very limited degree, which renders the value system useless as internal supervisor of adaptive change.

Let us start our discussion by remembering the neural structures whose activity corresponds to salient events. From the presented results, two possible ways to interpret such structures result: a.) They could be embodied structures, integrated in a sensorimotor context, whose meaning has to be investigated and interpreted within this context and during situated interaction with an environment. b) They could be value systems that autonomously perform judgments about the significance of a situation and rewire the agent accordingly.

The presented results hopefully illustrate how these two options exclude each other: An “embodied value system” is a *contradictio in adjecto*. The existence of reciprocal causal links between value system and behaviour generating systems causes semantic drift of the value signal, which results in anarchy of development (see Sect. 4.4). But how could a value system not be embodied? Surely, we do not want to introduce magic meaning sensors or a magic master value system that ensures that the other value systems work smoothly. This smells too much of what Rutkowska calls “[b]uck passing to evolution” [17]. If we struggle to explain the simple case without such scaffolding, the more abstract case will surely not become easier. The only way a value system architecture can work is a full embracement of the functional separation and pre-specification of meaning.

In the area of robotics, as shown in [22], we can design experiments rigidly enough to fixate meaning. But for an approach that aims at advancing past the stage of pre-specified motor programs, that refers to variable biomechanical properties in living organisms, the introduction of parts of the organism that are exempted from ontogeny, despite the constant material flux an organism undergoes, seems like a step backwards. It appears so inevitable that a random change would slightly change the context in which a value system is embedded, and the value-agnostic remainder of the organism would be unable to detect it

or do anything about it. Furthermore, both in the area of biological modelling and in robotics, there is another unpleasant side-effect resulting from the introduction of disembodied and non-adaptive value systems: The impossibility of novel values. A rigid structure with *a priori* meaning can only work in situations that rely on phylogenetic constancies, the generation of new values in situations that our ancestors could not even have dreamt of asks for a different explanation.

We do not want to question that structures like the ones described as value systems exist in living organisms and that they play an important role in the adaptation of behaviour. In contrary, we think that the investigation of such mechanisms is important and intriguing. We plan follow-up experiments to the ones presented in Sect. 4.4, to investigate possible embodied functions that “neural value structures” could have for the adaptation of behaviour². However, what we do want to question is that such components are or could be the loci of meaning. We question the idea that the generation of meaning can be separated functionally. Such components form part of an integrated system and their functionality both constrains and is constrained by this system they form part of, and therefore, they have to be interpreted as parts of a complex mechanism, not as encapsulated generators of judgment.

7 Conclusion

This paper does not have to be seen exclusively as a criticism of the value system as a locus of judgment, but as a general conceptual argument about correlation of neural activity with functional aspects of behaviour and how it does not entail, or even justify, the reduction of the respective function to the respective brain structure. Even though this point is not exactly novel, the enthusiasm with which researchers sympathetic to the embodied approach implement and develop “value system architectures”, in which a disembodied module is introduced to provide *a priori* specified criteria to guide embodied and situated lifetime development, provoked us to conduct the presented series of simple simulation experiments. These experiments illustrate the impossibility to reconcile functional reduction and the embodiment and situatedness of behaviour, which has been discussed in detail for the case of value system architectures, but extends to all models that feature a functional and structural separation of mechanisms of meaning generation from mechanisms of behaviour generation, i.e. all hybrid symbolic/embodied approaches to adaptive and intelligent behaviour: If a full-blown ghost in the machine has difficulties dealing with the variability of the external world, why would a vestigial ghost in the machine not face the same difficulties dealing with the variability of its bodily environment?

²A crucial aspect to change is a task that requires long term adaptive modulation of behaviour, which was neither the case nor necessary in this paper.

References

- [1] whoever showed that the baldwin effect can work
- [2] Barto: reinforcement learning in the basal ganglia 1995
- [3] Beer, R. D.: *Toward the Evolution of Dynamical Neural Networks for Minimally Cognitive Behavior*. In: P. Maes, M. J. Mataric, J.-A. Meyer, J. B. Pollack & S. W. Wilson (eds.): *From Animals to Animats 4. Proc. 4th Int. Conf. on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press 1996. 421–429.
- [4] Brooks, R. A., *Intelligence Without Reason*. Proceedings of 12th Int. Joint Conf. on Artificial Intelligence, Sydney, Australia, August 1991. pp. 569–595.
- [5] Edelman, G.: *The Remembered Present. A Biological Theory of Consciousness*. Basic Books, New York 1989.
- [6] Edelman, G.: *Neural Darwinism. The Theory of Neuronal Group Selection*. Oxford University Press, 1989.
- [7] Edelman, G.: *Naturalizing consciousness: a theoretical framework*. Proc Natl Acad Sci USA. Apr 29;100(9) 2003. pp 5520-5524.
- [8] Harnad Symbol grounding problem
- [9] Izquierdo–Torres, E. and I. Harvey:
- [10] Kandel, Eric R., James H. Schwartz and Thomas M. Jessel (eds.): *Principles of Neural Science*. Fourth Edition. McGraw & Hill, New York 2000.
- [11] Littmann Evolutionary reinforcement learning.
- [12] Nolfi, S and D. Floreano: *Evolutionary Robotics. The Biology, Intelligence, and Technology of Self-Organizing Machines*. MIT Press, Cambridge MA 2000.
- [13] Oyama, S.: *The Ontogeny of Information*.
- [14] Pfeifer, R. and C. Scheier: *Understanding Intelligence*. MIT Press, Cambridge MA 1999.
- [15] Rumelhart hinton backprop
- [16] Russel and Norvig.
- [17] Rutkowska, J.: *What's value worth? Constraining Unsupervised Behaviour Acquisition*. In: Proc. of the Fourth European Conference on Artificial Life 1997. 290–298.

- [18] Sporns, O., and G.M. Edelman: *Solving Bernstein's Problem: A Proposal for the Development of Coordinated Movement by Selection*. Child Dev. 64 (1993) 960–981.
- [19] Suzuki, M., Floreano, D. and Di Paolo, E. A., (2005). The contributions of active body movement to visual development in evolutionary robots. Neural Networks. 18(5/6) pp. 657-666.
- [20] Todd, M. and J. Miller:
- [21] Tuci, E., Quinn, M. and Harvey, I.: *Evolving fixed-weight networks for learning robots*. In Congress on Evolutionary Computation: CEC2002, IEEE Press 2002. pp 1970-1975.
- [22] Verschure, P., J. Wray, O. Sporns, G. Tononi and G.M. Edelman: *Multilevel analysis of classical conditioning in a behaving real world artifact*. Robotics and Autonomous Systems, 16 1995. 247-265.
- [23] Wittgenstein, L.: *Philosophical Investigations*.
- [24] Yamauchi, B. and Beer, R.D.: *Sequential behavior and learning in evolved dynamical neural networks*. Adaptive Behavior 2 (3) 1994:219-246.