A neural process model of learning to sequentially organize and activate pre-reaches

Jan Tekülve, Stephan K. U. Zibner, and Gregor Schöner*

Abstract— **Inspired by the longitudinal data of von Hofsten [1], we provide a neural process model of autonomously learning to direct pre-reaches toward visual objects. We build on an earlier neural dynamics account of pre-reaching [2], in which the elementary behaviors of visual fixation, reaching toward targets, returning to a resting position, closing, and opening the hand are tied to perceptual inputs and linked to a modeled muscle and effector system. In the current extension, the coupling of these elementary behaviors and their task-related recruitment emerge from an autonomous learning process that discovers which elementary behaviors in which sequential order are associated with success. The learning dynamics combines a memory trace of recent activation of behaviors and sequences with a neural representation of the reward that is inherent in moving the hand close to a visual object. We address how the temporally discrete reward events may be integrated into the time-continuous neural and learning dynamics. The simulated robotic model accounts for the three phases of activation, suppression, and re-emergence in the development of pre-reaching that were empirically observed by von Hofsten. These are attributed to the development of the sequential organization of movement and the stabilization of movement activation according to the spatial precision hypothesis.**

I. Introduction

Learning to reach and grasp objects in a baby's vicinity is a major developmental milestone that draws on a set of perceptual and motor skills. The longitudinal study of von Hofsten [1] assessed pre-reaching by recording visual fixation, arm movements, and hand configuration while infants observed stationary or moving objects. Von Hofsten identified three developmental phases. Infants of up to four weeks of age generated some movements of the hand in the general direction of the object, called pre-reaches, even though they often did not visually fixate the object. The number of pre-reaches dropped in a second phase between four and ten weeks of age, while the total time spent visually fixating the object increased. Starting around ten weeks of age, pre-reaching movements re-emerged with increasing frequency and often accompanied by visual fixation of the object. Von Hofsten also described a developmental trajectory for hand configuration.

Previously, we used the developmental signatures provided by von Hofsten to assess a neural processing account for the different developmental stages of pre-reaching [2]. Neural processes are captured at the level of the dynamics of neural population activation representing movement targets, and initial hand positions. The dynamics of neural oscillators generate timing signals that drive a muscle model and generate the physical movement. Visual fixation was modeled by a neural dynamics of covert attention. Initiation and termination of these different behaviors was accounted for by the neural dynamics of a set of activation variables. Their dynamic instabilities trigger the transitions among the different neural states that are entailed in reaching and looking.

We mapped the three developmental phases onto different dynamic regimes captured by three sets of values for model parameters [2]. Specifically, the connectivity among behaviors was changed from an early stage in which behaviors were not coupled and neural interaction within the neural dynamics was relatively weak to a more developed stage in which a specific pattern of coupling among behaviors was imposed and neural interaction was relatively stronger (the latter being consistent with the spatial precision hypothesis [3]). This mapping accounted for the developmental signatures observed by von Hofsten.

In this paper, we provide an account for the developmental process itself. We focus on the emergence of sequential organization of the comprised behaviors as an explanation for the three developmental phases found by von Hofsten's experiments. This complements earlier work on the sensorimotor aspects of learning the reaching component itself [4], [5]. Processes of autonomous learning from experience take the model through the developmental phases. Specifically, two sets of neural connections are learned. First, the coupling among behaviors emerges from an initial state without coupling to a coupling structure that reflects the intrinsic structure of the reaching task. Second, each behavior evolves from a less stable to a more strongly stabilized regime. Learning is instantiated by a combination of two factors. The dynamics of a memory trace represents the recent history of activation of each behavior. This memory trace is transformed into synaptic strength whenever a pre-reach was successful as indicated by an intrinsic reward signal.

The model learns as it autonomously generates sequences of motor acts through its time-continuous neural dynamics. We address the problem of how a reward signal that arises at discrete times at the end of a successful sequence of actions may impact on the time-continuous neural processes of learning. We show that the developmental trajectory that emerges from the learning process qualitatively matches von

Support by the EU project NeuralDynamics and by the Studienstiftung des Deutschen Volkes is gratefully acknowledged. We thank Sebastian Schneegans for suggestions about mechanisms of autonomous learning.

*JT, SKUZ, and GS are with the Institut für Neuroinformatik, Ruhr-Universität Bochum, Universitätsstr. 150, 44780 Bochum, Germany. {jan.tekuelve, stephan.zibner, gregor.schoener}@ini.rub.de

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016

Hofsten's developmental phases.

II. Methods

Dynamic Field Theory (DFT), a branch of neural dynamics, is a theoretical framework that provides a mathematically explicit way to model the evolution in time of neural population activity [6]. The main building blocks of DFT are dynamic neural fields (DFs) and dynamic neural nodes, both implemented as dynamical systems. Multidimensional dynamic neural fields model the activity of neural populations that are sensitive to certain common features. Locations within the field encode the corresponding value along the feature dimensions of the field. Feature dimensions may encode low-level sensor spaces like retinal location (to represent salient objects in the visual array, for instance), or also motor dimensions such as the Cartesian position of the hand in space. Peaks of activation represent the value along a feature dimension that is encoded at their location. Neural interaction within the dynamics of neural fields make such peaks attractor states. They may arise in dynamic instabilities driven by input.

A. Dynamic Fields and Nodes

The activation pattern, $u(\boldsymbol{x},t)$, of a dynamic neural field spanned over a N-dimensional feature space $\boldsymbol{x}$ evolves in time according to the integro-differential equation [7]:

$$\tau\dot{u}(\boldsymbol{x},t) = -u(\boldsymbol{x},t) + h + \sum_i s_i(\boldsymbol{x},t) + [w_{u,u} * \sigma(u)](\boldsymbol{x},t) \quad (1)$$

with

$$\sigma(u(\boldsymbol{x},t)) = 0.5\left(1 + \frac{\beta u(\boldsymbol{x},t)}{1+\beta|u(\boldsymbol{x},t)|}\right). \quad (2)$$

Here, τ defines the time scale of the neural dynamics together with the stabilization term, $-u(\cdot)$. The resting level, $h < 0$, of the field is negative. Inputs, $s_i(\boldsymbol{x},t)$, are summed. Localized patterns of input may drive associated locations towards a detection threshold that is set by the non-linear sigmoidal function $\sigma(\cdot)$. Once activation exceeds threshold, neural interaction is engaged. Formally, the thresholded activation pattern is convolved with an interaction kernel $w_{u,u}$, that consists of local excitation, and mid-range to global inhibition. Local excitatory coupling stabilizes peaks against decay, while lateral inhibitory coupling prevents activation from spreading out along the feature dimensions. Global inhibitory interaction may suppress activation at other field locations, effectively implementing a selection mechanism. With weak global inhibition, dynamic fields may contain multiple peaks at the same time.

Dynamic neural nodes

$$\tau\dot{u}(t) = -u(t) + h + c_{u,u}\sigma(u(t)) + \sum_i s_i(t), \quad (3)$$

have the same neural dynamics as fields, including a detection decision that is self-stabilized through non-linear local excitatory interaction with weight $c_{u,u}$. As representations, nodes are categorical in nature and can be understood as zero-dimensional fields ($N = 0$).

Fields may be combined with other fields, with sensor and motor systems, and with other building blocks into DFT architectures. Passing its activation through a sigmoidal threshold function, $\sigma(u(t))$, a field or a node may project onto other fields or nodes as excitatory or inhibitory input. Such projections may entail contraction or expansion along feature dimensions when the coupled field differs in dimensionality (for a detailed discussion, see [8]). A special case is a *switchable boost*, a neural node that projects homogeneously to a target DF as a boost of the field's resting level. Such a boost may push sub-threshold activation in the field through the detection instability and thus instantiate such sub-threshold activation as a self-stabilized peak. Conversely, a *peak detector* is a neural node that receives the summed supra-threshold activation of a DF and is tuned to go through the detection instability when there is a least one self-stabilized peak in the DF.

Another DFT building block is the memory trace. Here we use it for nodes using a learning dynamics that has two different time scales, $\tau_+ < \tau_-$,

$$\begin{aligned}\dot{\nu}(t) &= \frac{1}{\tau_+}(-\nu(t) + \sigma(u(t)))\sigma(u(t)) \\ &+ \frac{1}{\tau_-}(-\nu(t)(1-\sigma(u(t)))).\end{aligned} \quad (4)$$

The memory trace builds up whenever an activation variable, $u(t)$, is above threshold. It decays more slowly than it builds up when the activation variable falls below threshold.

B. Behavioral organization

Typically, a task is realized by a dynamic field architecture by activating a set of sub-tasks in a temporally organized fashion. For example, the task to pick up a bottle entails reaching, the transport of the hand toward the bottle, and grasping, closing the hand to grip the bottle. Such behavioral or process organization takes place in DFT by activating or deactivating elementary behaviors (EB) [9]. An EB is activated by driving its intention node through the detection instability. It may be deactivated, when its Condition-of-Satisfaction node (CoS node) becomes active. The intention-node works as a *switchable boost* for down-stream fields or nodes, which it may in turn activate through a detection instability. These down-stream structures may ultimately induce a representational or behavioral change, which is predicted by the intention node and preactivates an associated CoS field or node. When sensory input or internal feedback matches the prediction, the CoS becomes active. Internal feedback about achieving the goal of an intentional state may take the form of a *peak-detector*, which may then induces the transition to a new EB.

In the model, EBs receive input from a common task node that boosts their intention nodes. This is a place holder for intention nodes at a higher hierarchical level. We assume that an EB may be invoked in multiple, different tasks so that selection of preactivated EBs by a task node may be

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016

one form of behavioral organization. On the other hand, interaction among EBs within a single task is mediated by dynamical "pre-condition" nodes that are co-activated by the corresponding task node (see Figure 1 for an explanation).

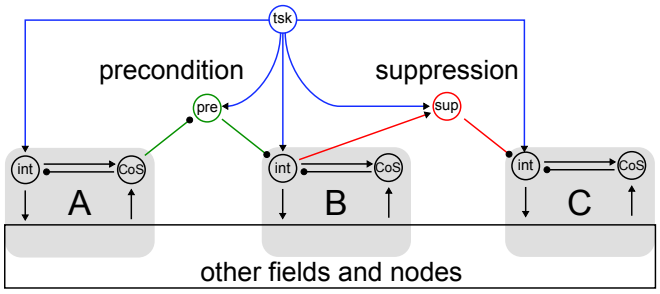

Fig. 1. This figure displays three elementary behaviors A, B and C connected through behavioral organization. A and B are connected through a precondition node (green) that prevents B from activating unless the CoS node from A is above threshold. B suppresses C through a suppression node (red) that prevents any activation of C while B is active. All EBs and the behavioral organization nodes are activated through a common task node (blue).

C. Spontaneous behavioral activation

In our developmental account, we assume that early in development, both forms of behavioral organization, preactivation of elementary behaviors by the task node and activation of the inhibitory pre-condition nodes, are not articulated. EBs may therefore spontaneously activate or deactivate. In the model, this happens because the intentional neural node has generic activation-inhibition neural dynamics that may enter an oscillatory regime, in which activation drives inhibition, which suppresses activation, and then decays itself, enabling activation to rise again [7]. Earlier, Perone and Spencer [10] suggested that such neural oscillation may be a generic mechanism for exploratory behavior in DFT. A periodically activated intention node may still be inhibited through forms of behavioral organization. It may also be lifted out of the oscillatory regime when it receives a sufficient boost from the task node as happens during learning (see Figure 2 for a sketch). A typical time course of activation, $u(t)$,

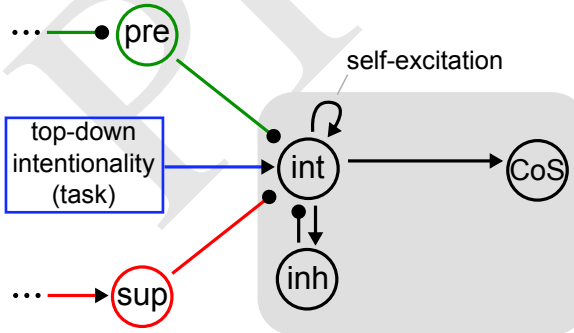

Fig. 2. This figures is a sketch of a spontaneously activating and deactivating intention node that acts as a neural oscillator modulated by behavioral organization and task input.

and the corresponding thresholded activation, $\sigma(u(t))$, of an oscillatory intention node is shown in Figure 3.

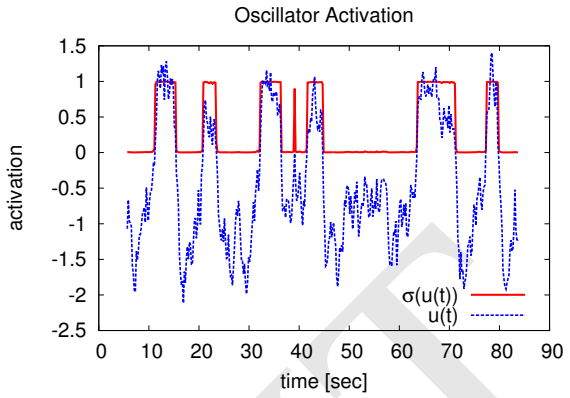

Fig. 3. This figure shows an exemplary activation time course, $u(t)$, of an oscillatory intention node. Only activation above threshold, $\sigma(u(t))$, activates downstream behaviors.

III. Model

In this paper we build on phase one of the pre-reaching architecture that was previously presented [2] (see Figure 4) as the initial condition of a developmental process model. The five elementary behaviors *fixate*, *reach*, *rest*, *open*, and *close* (Figure 5) are initially in the dynamic regime of spontaneous behavior activation (or "behavioral babbling") and are not initially coupled. We first explain the components of the architecture and then discuss the learning process.

A. Elementary Behaviors

The *fixation* behavior operates on a dynamic neural field defined over the visual array in retinal coordinates. This fixation field receives a salience pattern as input from the simulated vision sensor. When the fixation behavior is activated, its intention node boosts the resting level of the fixation field. As a result, saliency input from the visual array may then induce through a detection instability, a single peak of activation at the visual location with most salience. The intentional node of the fixation EB is self-stabilized and drives its inhibitory component, forming a neural oscillator. A similar notion has been used previously to account for patterns of infant looking [10]. A peak detector node activates the CoS node when a peak builds in the fixating field. A peak in the fixation field projects onto the reaching field as described by a coordinate transform from visual to Cartesian workspace. In this simplest model, the transformation is assumed given and trivialized by keeping the head and camera orientation fixed. The fixation behavior thus takes the form of covert attention to a salient visual location, which determines the location to which a reach is directed.

The *reaching* behavior is similarly controlled by an intention node that is initially in an oscillatory regime and globally boosts the reaching field, inducing a single peak of activation that represents the current reaching target in Cartesian work space. In the presence of a peak in the fixation field, the reaching field most likely selects the fixated (covertly attended) visual location as a reaching target. In

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016

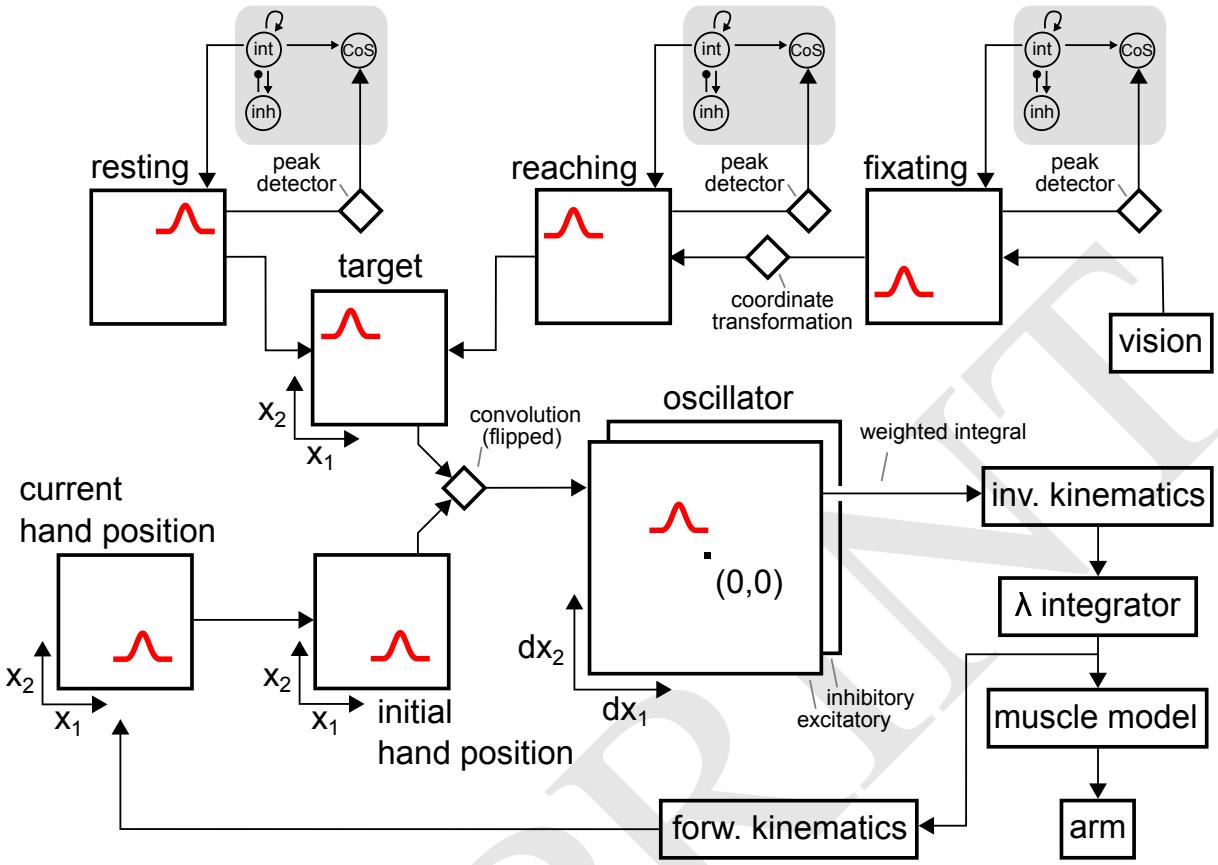

Fig. 4. This figure shows a schematic overview of how the three elementary behaviors 'reach', 'rest' and 'fixate' couple into movement generation. Squares with red activation peaks denote fields. Diamonds along connections are indicators of reference frame transformations. All three behaviors are driven by oscillatory intention nodes resulting in spontaneous forming and vanishing of peaks in the corresponding fields. Without a peak in the fixating-field influencing the reaching-field, peaks in the reaching-field emerge at a random position. Peaks in the fixating-field emerge at the position of the visual stimulus and peaks in the resting peak always form at the resting position. Once a peak is present in the target-field the movement generation architecture drives the arm to the corresponding workspace position. Nodes and connections realizing behavioral organization aside from the oscillatory intention nodes, both within each EB and between them, are not shown in this figure. For more details, please refer to [11]. Note that a similar network of fields translates the targets of the elementary behaviors 'open' and 'close' into movements of the hand.

the absence of a fixation peak, the reaching target arises as a fluctuation induced peak at a random location. Neural noise of considerable strength is present in all fields. A peak in the reaching field triggers a chain of events in the movement generation architecture (introduced in more detail in Zibner and colleagues [11]). The direction and extent to the movement are determined by transforming the reaching target into a coordinate frame that is centered on the initial hand (endeffector, eef) position. This coordinate transform is neurally implemented by convolving the reaching field with a neural field representing the initial hand position.

Triggered by input from the transformed reaching field, a two-layer field of neural oscillators generates timing signals in approximately Gaussian shape. The outputs of the different oscillators are combined with a learned set of weights that transform into virtual end-effector velocity, whose peak velocity and duration ultimately move the hand to the desired position. The velocity command is transformed into joint-space using an inverse kinematic map and is subsequently path-integrated to form an internal representation of the desired joint configuration. The resulting virtual trajectory of joint vector drives a set of muscles, modeled in simplified form as critically damped harmonic oscillators, by shifting their resting lengths. The torques predicted by the muscle models drive the robot arm's dynamics. The virtual trajectory is also used in the form of a corollary discharge to predict the hand position in space based on a forward kinematic model.

The *resting* behavior is an alternate source of specification of a movement target for the arm. An activation of the resting intention node always leads to a peak forming in the resting field at a default resting position. That peak projects onto the target field, where it converges with input from the reaching field. When input from the two sources of specification induces two peaks in the target field, then the eef will move toward an averaged position.

The behaviors *open* and *close* similarly converge on shared neural processing structure that generates hand movement and is similar to the architecture for reaching of Figure 4. Both behaviors generate movement with default movement parameter values along a single dimension, the hand opening. These are represented in associated fields and activated by oscillatory intention nodes.

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016

B. Learning Framework

We assume that a pattern of coupling among the five elementary behaviors exists prior to learning. This pattern is illustrated in Figure 5 and is meant to capture in a qualitative way the functional role of the basal ganglia. Each elementary behavior inhibits four precondition nodes, which project inhibitorily onto the four other elementary behaviors. This coupling structure provides for the potential of sequentiality: An activated precondition node prevents its target elementary behavior from becoming activated. Only when the elementary behavior (the "precondition") that inhibits the precondition node has been performed, is the target elementary behavior released from inhibition and may become activated. Completion of the precondition behavior is signaled by the Condition-of-Satisfaction of that behavior (see Figure 6 for the fine structure of this coupling structure).

The precondition nodes are activated by input ($s^{\mathrm{pre}}_{\mathrm{AtoB}}$ in Figure 6) from a task node that effectively selects the set of sequential constraints relevant to the task. Early in development, we assume this set of connections has zero strength, so that no sequential structure is active. During learning, the sequential order of activation of elementary behaviors that leads to reward strengthens the projections from the task node to the appropriate precondition nodes, creating a task network of sequential constraints.

A second substrate for learning is a set of projections from the task node to each elementary behavior ($s^{\mathrm{eb}}_{\mathrm{A}}$ in Figure 6). This network primes elementary behaviors engaged in the task, making them easier to activate and their activated state more stable. Early in development, we assume this second set of connections also has zero strength so that all elementary behaviors are equally easy to activate. During learning, task input to the set of elementary behaviors that leads to reward is strengthened.

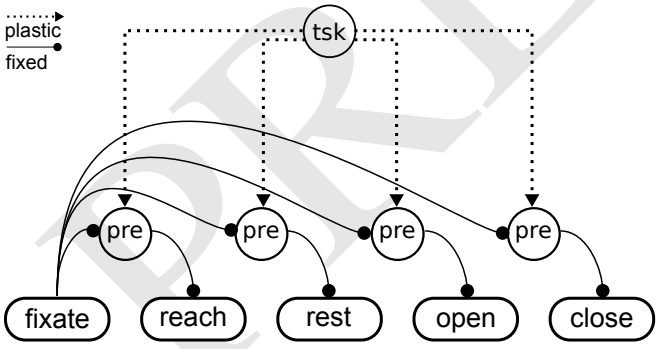

Fig. 5. The four precondition connections outgoing from the fixate behavior. Analog connections exist for the other four behaviors resulting in a total of twenty preconditions. The precondition nodes need sufficient input from the task node in order to prevent the corresponding behavior from activating. All task-to-precondition connections are not sufficient initially and their connection strength is subject to learning.

The neural architecture spontaneously generates behavior as elementary behaviors are activated through the inherently oscillatory nature of their neural dynamics. As this happens, the system learns based on two factors. (1) Memory traces that reflect the recent behavioral history and its sequential order, and (2) reward events that occur when a pre-reach brings the hand sufficiently close to a visually perceived object. We briefly review the two factors and explain how the temporally discrete onsets of the reward signal are integrated into the time-continuous activation and learning dynamics.

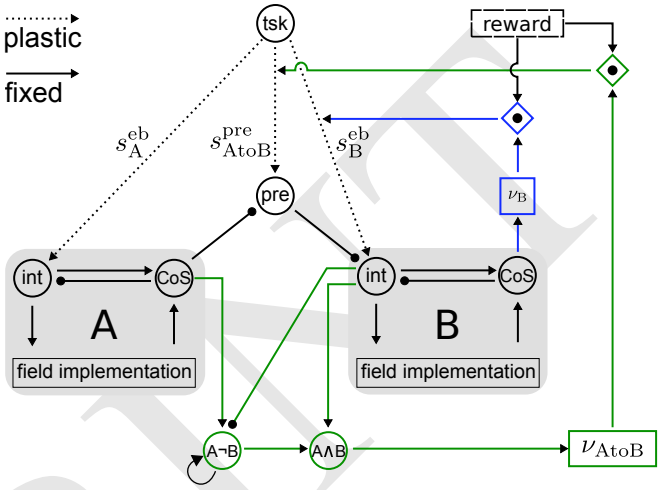

Fig. 6. Two Elementary Behaviors A and B and their corresponding intention and CoS nodes. The plastic connections to the task node are denoted with dotted lines. Plastic connections are adapted depending on a global reward signal and the corresponding memory trace activation ν_{B} or ν_{AtoB}. The part responsible for the memory trace acquisition for the precondition connection is shown in green, while memory trace acquisition shown for a single behavior (B) is shown in blue. Memory trace acquisition for behavior A or the sequence BA are analog to the depicted mechanisms and not shown here.

1) Memory traces: The CoS node of any individual elementary behavior, B, builds a memory trace, ν_{B}, according to Equation 4. This memory trace represents the recent activation of behavior B even after the intention and CoS activation of B have subsided, and thus helps carry the activation history into the time interval during which reward is signaled.

To build a similar representation of the sequential structure of recent activation patterns we use a pair of neural nodes, 'A¬B' and 'A∧B' (see Figure 6) for each pairing of two elementary behaviors, A and B. Their connectivity is inspired by the serial encoder network proposed as a model of the basal ganglia by Houk and colleagues [12]. The 'A¬B' node receives excitatory input from the CoS node of behavior A and inhibitory input from the intention node of behavior B. This node is therefore activated if A has been performed in the absence of ongoing activation of B. Once activated, the 'A¬B' node is stabilized by recurrent self-excitation. The 'A∧B' node receives excitatory input from both the 'A¬B' node and from B's intention node. This node is therefore activated only if previously behavior A has been performed in the absence of B, but B is now active. This nodes thus flags the sequential nature of the activation of first A, then B. According to Equation 4, the memory trace, ν_{AtoB} keeps a record of this sequentiality detection even after the individual intention and CoS nodes have become deactivated.

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016

2) Reward signal: A reward signal, $r(t)$, is generated autonomously, whenever the hand moves sufficiently close to the visual target. This happens through a neural field, u_{rew}, in which a peak is induced when the hand and the visual target overlap sufficiently. Learning occurs in a well-defined time window after the onset of such reward detection so that it does not depend on how long the hand remains near the target. This translation of discrete events into time-continuous signals is a core problem in autonomous learning with neural dynamics. It is solved here through an eligibility trace that is generated by a pair of excitatory, u_{exc}, and inhibitory,u_{inh}, activation variables, both of which receive input, $\sigma(u_{\mathrm{rew}})$, from the reward field:

$$\tau_{\mathrm{exc}}\dot{u}_{\mathrm{exc}}(t) = -u_{\mathrm{exc}}(t) + h + \sigma(u_{\mathrm{rew}}) - c_{u_{\mathrm{inh}},u_{\mathrm{exc}}}\sigma(u_{\mathrm{inh}}(t)) \quad (5)$$

$$\tau_{\mathrm{inh}}\dot{u}_{\mathrm{inh}}(t) = -u_{\mathrm{inh}}(t) + h + \sigma(u_{\mathrm{rew}}) \quad \text{with } \tau_{\mathrm{exc}} < \tau_{\mathrm{inh}}. \quad (6)$$

When that input becomes non-zero, both neurons go through the detection threshold, but excitation does so earlier than inhibition because its relaxation time is faster ($\tau_{\mathrm{exc}} < \tau_{\mathrm{inh}}$). Once the inhibitory node goes above threshold, it deactivates the excitatory variable again, which therefore only generates a single transient pulse of activation. Passed through the threshold function, this pulse, $\sigma(u_{\mathrm{exc}}(t))$, serves as the transient reward signal $r(t)$. Once the hand moves away again from the target, the peak in the reward field decays, the transient system re-levels its activation pattern without generating another excitatory pulse, and the system is ready for a potential new reward.

3) Learning rule: Both a memory trace and a reward signal must be present, for any of the projections to be strengthened. We use this simplest form of a learning rule:

$$\dot{s}^{\mathrm{eb}}_{\mathrm{A}}(t) = r(t)\lambda^{eb}\nu_{\mathrm{A}} + (1 - H(\nu_{\mathrm{A}}))r(t)\gamma^{eb} \quad (7)$$

$$\dot{s}^{\mathrm{pre}}_{\mathrm{AtoB}}(t) = r(t)\lambda^{pre}\nu_{\mathrm{AtoB}} + (1 - H(\nu_{\mathrm{AtoB}}))r(t)\gamma^{pre} \quad (8)$$

where $H(\cdot)$ is the Heaviside step function, λ^{eb} and $\lambda^{pre} > 0$ are the respective learning rates, and γ^{eb} and $\gamma^{pre} < 0$ the corresponding rates of unlearning. The idea is that only while a reward signal is detected, those connection strengths, s, are strengthened who reflect recently active elementary behaviors or recently observed sequential activation patterns. All other strengths are weakened.

IV. Experiments

We implemented and simulated the autonomous learning dynamics of pre-reaching in the previously used robotic scenario [2]. The simulated Nao robot operated in a $200\,\mathrm{mm} \times 200\,\mathrm{mm}$ workspace containing the resting position and two possible target positions (see Figure 7). A single simulated reaching target was presented to the visual system at all times during the experiment. The global reward signal was triggered when movement ended with the eef within a radius of $15\,\mathrm{mm}$ of the simulated target position. Once a reward was emitted, the simulated target was switched to the other possible target position.

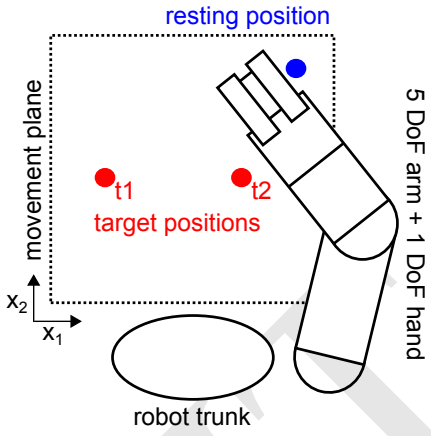

Fig. 7. This sketch depicts the experimental setup, showing the robots reaching workspace ($200\,\mathrm{mm} \times 200\,\mathrm{mm}$), the resting position and the two alternating target positions.

Two discrete target positions were presented rather than a continuously moving target, as this made it easier to evaluate the reaching behavior. At all times, behaviors were autonomously generated by the architecture based on spontaneous activations of the intention nodes. In the present simulation, *open* and *close* behaviors had no effect on the reward signal. This allowed us to contrast behavior that is not expected to be affected by learning with behavior that is affected. The goal was to learn the precondition connectivity as well as to learn the connectivity from task node to the elementary behaviors for *fixate* and *reach*, while finding *open* and *close* to remain invariant.

The learning and unlearning rates were set to $\lambda^{\mathrm{pre}} = -\gamma^{\mathrm{pre}}$ for the preconditions and to $\lambda^{\mathrm{eb}} = -\gamma^{\mathrm{eb}}$ for the elementary behaviors. We concluded from the data of von Hofsten [1] that the precondition should be learned faster ($\lambda^{\mathrm{pre}} > \lambda^{\mathrm{eb}}$) than the elementary behaviors, because sequentiality is already established to some extent in phase 2 and performance continues to improve in later phases.

We ran the experiment three times for a simulated time of 26 hours with a single reach from the resting position to target $t1$ (see Figure 7) taking $8.5\,\mathrm{seconds}$. The proposed learning dynamics succeeded in establishing the precondition between *fixate* and *reach* as well as the task connections to the elementary behaviors *fixate* and *reach* (see Figure 8 and Figure 9). We also measured the number of movements every hour and observed a U-shape over learning time, similar to von Hofsten's experimental observation (see Figure 10).

Thus, the three different phases of pre-reaching develop autonomously within the same neural architecture. For each phase, characteristic movement trajectories are depicted in Figure 11. The first phase is dominated by movements to random locations that occasionally hit target positions. In the second phase, the overall number of pre-reaching movements decreases, as the emerging precondition inhibits some reaching attempts. The percentage of movements directed towards the target position begins to outweigh random

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016

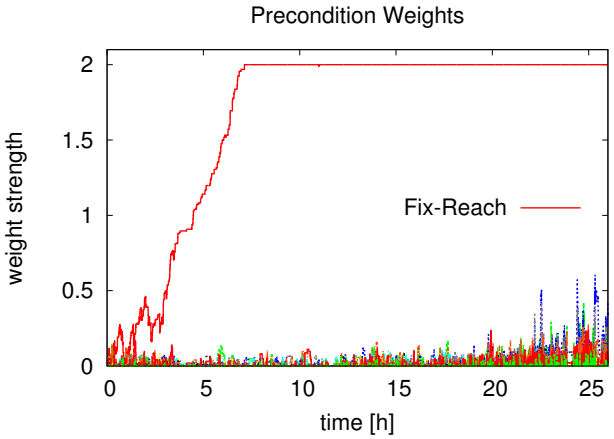

Fig. 8. This figure shows the development of weights of all twenty possible preconditions across 26 hours simulated time for one exemplary simulation run. The precondition *fixate* before *reach* is the only learned precondition, which reaches the maximum synapse weight of two. The other precondition weights are not able to pierce the threshold of one, because they are subject to the unlearning rule as they are not required for achieving the reward. Note that the weights tend to reach higher values in the last third of the simulation. This results from the higher activation rate of the learned elementary behaviors leading to an increased probability for activating the corresponding precondition memory traces.

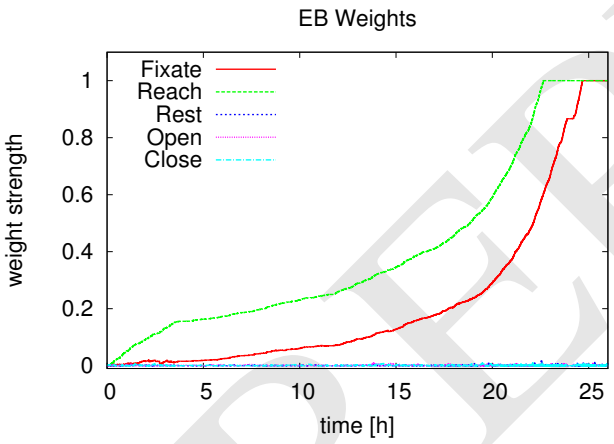

Fig. 9. This figure shows the development of weights of the five elementary behaviors across 26 hours simulated time for one exemplary simulation run. The weights from the task node to the elementary behaviors *fixate* and *reach* grow continuously while the behaviors *close hand*, *open hand* and *rest* do not. Note that the rate of change of *reach* decreases slightly around the time of the learned precondition (after around 4 hours), but then grows in an exponential fashion as more fixation and reaches lead to more reward, which in turn leads to an increase in connection strength. The *reach* behavior is learned faster than *fixate*, because it is temporally closer to the reward event, thus less memory trace is decayed once the learning rule is applied.

movements. Accuracy towards the targets and the overall number of reaches further increase in the third phase as the corresponding elementary behaviors become easier to activate and stabilize. Movements take place between the two target positions and the resting position.

V. Discussion and Conclusion

We have presented an account of autonomous learning that captures the developmental process of pre-reaching in

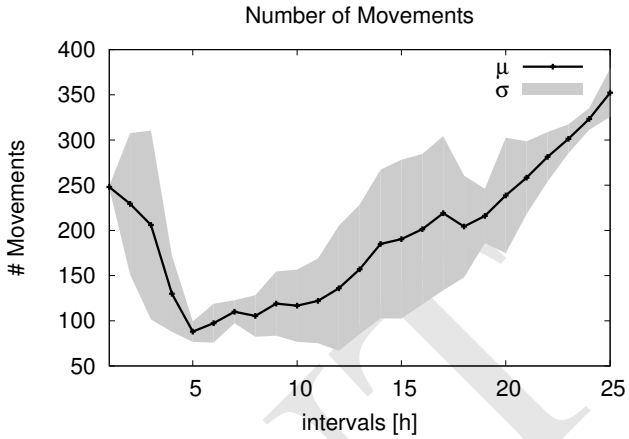

Fig. 10. This figure shows the number of movements elicited in each hour averaged over the three simulation runs, μ denoting the mean value and σ the standard deviation. The average number of movements per hour decreases with the establishment of the precondition around hour 5 and continues to rise again due to individual strengthening of relevant behaviors. Note that the standard deviation σ between the three simulations is lowest around the time of the learned precondition and after learning, while the learning process of the elementary behaviors has a higher variance across simulation runs.

infants. A dynamic field architecture provides elementary behaviors from salience-based visual input to the generation of motor commands that drive muscles and move the hand in space [2]. Early in development, behaviors are spontaneously activated based on the intrinsic bistability of the neural dynamics, but not sequentially organized.

Two neural connectivity patterns evolve during learning (see Figure 6). One set of connections controls the sequential activation of elementary behaviors. The learning process for this set of connections is mediated by a sequentiality detector, a neural structure inspired by models of learning in the basal ganglia [12]. The other set of connections in effect strengthens neural interaction consistent with the Spatial Precision Hypothesis [3]. Over learning, therefore, the neural dynamics becomes more strongly self-stabilizing and generates fewer uncoordinated spontaneous switches of activation.

The time-continuous learning rule for both connectivity patterns combines a memory trace of recent activation with an endogenous reinforcement signal. A role of reinforcement in learning to reach is supported by recent empirical findings [13]. Learning episodes are embedded in the continuous time neural dynamics through a transient reinforcement signal triggered each time the hand comes sufficiently close to the visual object. The observed U-shape is modelled by assuming that the learning of the connectivity pattern required for sequential organization is faster than the learning of the connectivity that strengthens interaction within each component.

The model does not yet include learning to open and close the hand. Such an extension would be possible by including a component to the reinforcement signal that is sensitive to the state of the hand.

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016

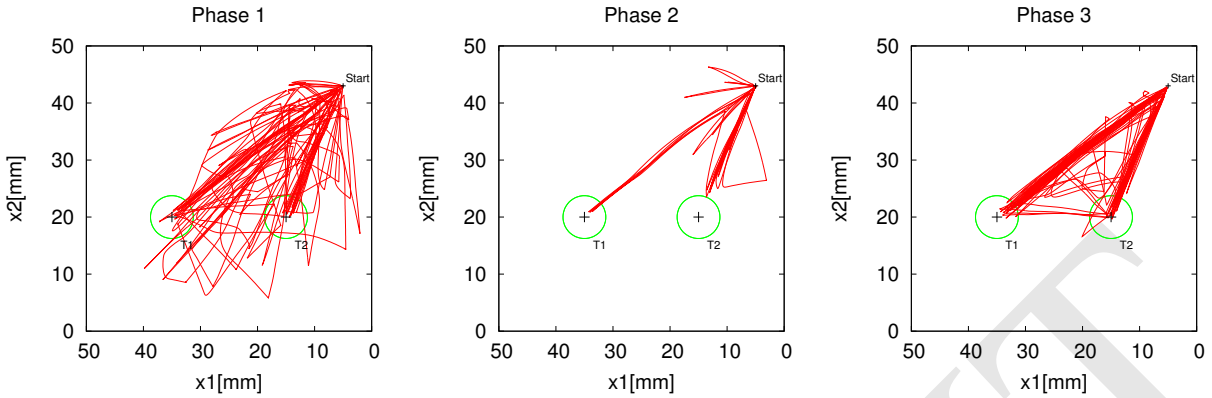

Fig. 11. These three figures show the eef trajectories elicited in the three developmental phases. The starting position is in the top right corner of the workspace and the two target areas are depicted with green circles. Each phase was plotted for 20 minutes of simulated time. In phase one targeted and non-targeted movements can be observed, while in phase two the number of overall movements decreases. In phase three targeted movements outweigh non-targeted movements and the overall number of movements increases again.

The notion of motor "babbling" has been employed in a number of projects that take inspiration from infant development [14]–[16]. Among them Shaw and colleagues [15] and Narioka and Steil [16] also dealt with modeling the U-shape found by von Hofsten, and achieved it by altering the parameters of the "babbling"-algorithm across different phases of development. A strength of the proposed learning processes is the autonomous nature of learning from experience. In fact, the initial exploratory activation of elementary behaviors transitions into coordinated and goal-directed behaviors based on the intrinsic bistability of the neural dynamics. It would be attractive to explore how engaging the system in multiple different tasks that involve the same elementary behaviors may lead to a form of hierarchical organization of the neural dynamics in which a specific task set may be selectively activated.

In our model, reaching emerges as the sequential organization of visual fixation followed by reach initiation is learned. This may appear to contradict recent empirical findings [17] according to which infants reach and then learn to look at the reaching location. The contradiction is only apparent, however, as the infants globally look at the object any time they reach. What they learn is to direct their gaze to the location on the object at which their hand makes contact. This fine structure of the reaching behavior is not yet addressed in our model. The data [17] provides important constraints for such future extensions of the model.

References

[1] C. von Hofsten, "Developmental changes in the organization of prereaching movements.," *Developmental Psychology*, vol. 20, no. 3, pp. 378–388, 1984.

[2] S. K. U. Zibner, J. Tekülve, and G. Schöner, "The Sequential Organization of Movement is Critical to the Development of Reaching: A Neural Dynamics Account," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2015 Joint IEEE International Conferences on*, pp. 39–46, 2015.

[3] A. R. Schutte, J. P. Spencer, and G. Schöner, "Testing the Dynamic Field Theory: Working Memory for Locations Becomes More Spatially Precise Over Development," *Child Development*, vol. 74, no. 5, pp. 1393–1417, 2003.

[4] N. E. Berthier, M. T. Rosenstein, and A. G. Barto, "Approximate optimal control as a model for motor learning.," *Psychological review*, vol. 112, no. 2, p. 329, 2005.

[5] D. Caligiore, D. Parisi, and G. Baldassarre, "Integrating reinforcement learning, equilibrium points, and minimum variance to understand the development of reaching: A computational model.," *Psychological review*, vol. 121, no. 3, p. 389, 2014.

[6] G. Schöner and J. Spencer, *Dynamic Thinking. A Primer on Dynamic Field Theory*. Oxford University Press, 2015.

[7] S.-i. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological cybernetics*, vol. 27, no. 2, pp. 77–87, 1977.

[8] S. K. U. Zibner and C. Faubel, "Dynamic scene representations and autonomous robotics," in *Dynamic Thinking. A Primer on Dynamic Field Theory*, ch. 9, pp. 227–245, Oxford University Press, 2015.

[9] M. Richter, Y. Sandamirskaya, and G. Schöner, "A robotic architecture for action selection and behavioral organization inspired by human cognition," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2457–2464, 2012.

[10] S. Perone and J. P. Spencer, "Autonomy in action: linking the act of looking to memory formation in infancy via dynamic neural fields," *Cognitive science*, vol. 37, no. 1, pp. 1–60, 2013.

[11] S. K. U. Zibner, J. Tekülve, and G. Schöner, "The neural dynamics of goal-directed arm movements: a developmental perspective," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2015 Joint IEEE International Conferences on*, 2015.

[12] J. Houk, C. Bastianen, D. Fansler, A. Fishbach, D. Fraser, P. Reber, S. Roy, and L. Simo, "Action selection and refinement in subcortical loops through basal ganglia and cerebellum," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1485, pp. 1573–1583, 2007.

[13] J. L. Williams and D. Corbetta, "Assessing the Impact of Movement Consequences on the Development of Early Reaching in Infancy," *Frontiers in Psychology*, vol. 7, no. April, pp. 1–15, 2016.

[14] R. Saegusa, G. Metta, G. Sandini, and S. Sakka, "Active motor babbling for sensorimotor learning," *IEEE International Conference on Robotics and Biomimetics*, pp. 794–799, 2008.

[15] P. Shaw, D. Lewkowicz, A. Giagkos, J. Law, S. Kumar, M. Lee, and Q. Shen, "Babybot challenge: Motor skills," in *Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2015 Joint IEEE International Conference on*, pp. 47–54, IEEE, 2015.

[16] K. Narioka and J. J. Steil, "U-shaped motor development emerges from goal babbling with intrinsic motor noise," in *Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2015 Joint IEEE International Conference on*, pp. 55–62, IEEE, 2015.

[17] D. Corbetta, S. L. Thurman, R. F. Wiener, Y. Guan, and J. L. Williams, "Mapping the feel of the arm with the sight of the object: on the embodied origins of infant reaching.," *Frontiers in psychology*, vol. 5, no. June, p. 576, 2014.

Published as: Jan Tekülve, Stephan K U Zibner, and Gregor Schöner. A neural process model of learning to sequentially organize and activate pre-reaches. In *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2016 Joint IEEE International Conferences on*, pages ??–??, 2016