Computational models of the emergence of self-exploration in 2-month-old infants

Josua Spisak*[1], Jan Benad*[2], Johannes Heidersberger[3], Stephan Verschoor[4], Pablo Lanillos[5], Dongheui Lee[3], Manfred Eppe[2], Stefan Wermter[1], Matej Hoffmann[6], and Sergiu Tcaci Popescu[6]

***Abstract*— Infants actively explore the relationship between actions and their associated effects (i.e., sensorimotor contingencies) before full-blown agency emerges. While there is experimental evidence for this development during the first year of life, the interplay of the associated cognitive processes is not yet well understood. This paper uses computational modeling to examine how exploratory behavior develops, based on one of the earliest experiments showing such behavior. In a seminal study of Rochat & Striano (1999), 2-month-old infants, contrary to newborns, showed differential behavioral patterns towards mouth-contingent sounds versus random sounds. This is interpreted as early evidence for action-effect exploration. We consider seven potential developmental factors as possibly explaining the emergence of active exploratory behavior in 2-month-olds: i) outcome prediction, ii) novelty preference, iii) fatigue, iv) strength, v) memory, vi) sensory noise, and vii) motor noise. These factors were implemented in both a supervised-learning model and a reinforcement learning model. Results from both models indicate that increased memory capacity with age is a key developmental factor underlying active exploration and, possibly, agency.**
Our code is published at: Computational models

I. Introduction

A sense of agency requires having a library of actions that are mapped to their respective effects. According to ideomotor theorizing [1], the ability to use such action-effect knowledge for anticipating action effects is the basis of voluntary action. Moreover, the sense of agency, or the feeling that one causes desired outcomes by acting, is often hypothesized to be a fundamental building block of the self [2]. Suffice it to say that the developing ability to map actions to their effects is a central precursor to a meaningful understanding of the world and (its relation to) the self (e.g. [3], [4], [5]).

In our study, we aim to investigate the developmental factors of the exploratory behavior in infants by utilizing computational modeling. We focus on the emergence of active exploration shown in Rochat & Striano [6]. Their study investigated early indicators of agency detection and voluntary action by comparing newborns' (0mo) and 2-month-olds' (2mo) oral activity on a pacifier that produced auditory feedback. If infants pressed the pacifier with their mouth above a certain threshold, a sound was played. In the analog condition, the frequency of the sound was proportional to the infant's mouth pressure, while in non-analog condition, the frequency of the sound was random. Rochat & Striano [6] showed that by two months of age, infants reacted differently to the analog condition, namely by pressing the pacifier more often around the pressure threshold that triggered the sound. This suggests that 2mo detect the difference between the conditions and learn to modulate their behavior in the analog condition whereas 0mo do not. One can argue if this truly shows voluntary action, an emerging sense of agency, or even causality detection [7], [8]. However, most would agree that the study shows sensorimotor contingency detection and, therefore, an important precursor to the sense of agency and voluntary action. Although this is an important finding giving insight into the developmental timeline of voluntary action, it remains unclear from the current data which exact underlying factors drive the measured developmental change.

Our computational modeling approach, by the implementation of theoretical sensorimotor mechanisms, attempts to determine the underlying developmental factors. Our model serves as a simulation to demonstrate whether the developmental trajectory described by Rochat & Striano [6] can be generated by parsimonious mechanisms (see [9] for a similar approach). Furthermore, by merging developmental and sensorimotor theory on the one hand and modeling approaches on the other, we hope to generate new theoretical insight into the development active exploration and agency.

Drawing from ideomotor theory, developmental and cognitive psychology, we identified seven potential underlying developmental factors, which are often used in modeling behaviors: i) outcome prediction, ii) novelty preference, iii) fatigue, iv) strength, v) memory, vi) sensory noise, and vii) motor noise. We grant that this list is not all-encompassing and that there is some overlap (e.g. no learning without memory),

*Equal contribution
[1]Knowledge Technology, Department of Informatics, University of Hamburg, Hamburg, Germany
[2]Institute for Data Science Foundations, Hamburg University of Technology, Hamburg, Germany
[3]Autonomous Systems Group, TU Wien, Vienna, Austria
[4]Cognitive Systems Lab, Faculty of Mathematics and Computer Science, Bremen University, Bremen, Germany
[5]Neuro AI and Robotics Group, Cajal Neuroscience Center, Spanish National Research Council, Madrid, Spain
[6]Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czechia

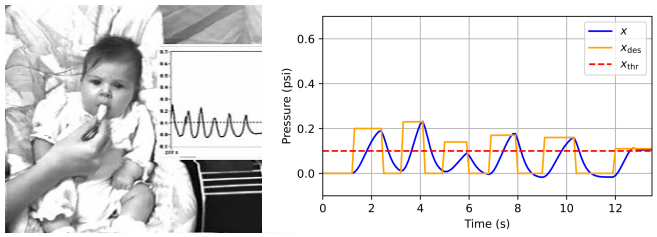
This collaboration and work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the Priority Programme "The Active Self" (SPP 2134), specifically through the DFG projects 402776968 (JS, JB, ME & SW), 402778716 (JH & DL), 467045002 (PL) and 467058220 (SV). We gratefully acknowledge funding by the Czech Science Foundation GA ČR project no. 25-18113S (MH, STP) and the project Mobility ČVUT MSCA-F-CZ-I, no. CZ.02.01.01/00/22_010/0003405 (STP).

nevertheless, these concepts are suitable for mathematical description and sufficiently distinguishable to model them. We will attempt to determine which developmental factors drive the developmental pattern found by Rochat & Striano [6] by calculating the correlation between the factors and the behavioral patterns found in [6]. All possible factor combinations are tested and those with the highest correlation with the behavioral patterns will be considered to model the underlying developmental process best. We will now briefly introduce all the developmental factors we considered.

Outcome prediction (i) in our model refers to the tendency to minimize the distance between the predicted pitch and the actual pitch. In action control terms, this refers to the process of action-effect learning. Action-effect knowledge is thought to develop during the first year of life, while full-blown voluntary action is thought to develop during the second half of the first year [7], [8]. Novelty preference (ii) refers to how much emphasis our models place on the next sound being different from the last. Its counterpart is familiarity preference. Although both phenomena have been extensively used to assess discrimination abilities in infancy [10], [11], it is unknown under what circumstances one or the other occurs [12]. Both are thought to accommodate learning. While we have not implemented familiarity preference, stimuli that one can predict can be considered familiar. Hence, action outcome prediction, partially, overlaps with familiarity.

We also incorporate fatigue (iii) and strength (iv). Fatigue, as captured by daily sleeping time, progressively decreases with infants age [13]. Furthermore, muscles quickly build strength during infancy [14]. Memory (v) is simulated in our models as the number of past sounds that are memorized. Evidence from the mobile paradigm suggests that memory retaining duration increases with age and recognition cues can decrease in specificity with age [15]. Sensory noise (vi) refers to how fine-grained the ability of the model is to distinguish between sounds. In order to detect contingencies between actions and sound, it is crucial to be able to distinguish the produced tones. There is evidence that 0mo can distinguish their mother's voice from others' [16] and show a preference for their native language [17], [18], suggesting sophisticated auditory analysis even at birth. However, to the best of our knowledge there are no studies that investigate the distinction of sine wave tones. Regarding sucking and swallowing behaviors, it has been suggested that this is the most advanced adaptive behavior a newborn infant has [19]. We modeled the development of this ability as a reduction of motor noise (vii) with age. The actual applied force in sucking corresponds to the intended force to which we add a motor noise.

The developmental factors were modeled in two neural network architectures differing in learning mechanism: Self-Supervised Learning (SSL), where the agent minimizes a cost function, and Reinforcement Learning (RL), where the agent maximizes the expected reward. While this sounds similiar the underlying algorithms beneath these purposes are quite different. Both architectures have been successfully used to model sensorimotor behavior [20], [21], [22], [23]. Our modeling approach and comparing its results with the main findings of Rochat & Striano [6] will allow us to shed light on the question: What developmental factors cause the developmental change leading to 2mo adapting to specific contingencies? Furthermore, comparing two architectures with different assumptions provides us with a compelling way to evaluate the plausibility of our findings. Convergent results will be seen as evidence while divergence may reflect different assumptions and uncertainty of the two approaches.



(a) Reference [6]. (b) Replication.

Fig. 1: Comparison between pressure dynamics generated by infants and model. (a) A picture of an infant and the dynamics of the pressure applied to the pacifier by the infant, as reported in Rochat & Striano [6]. (b) The biomechanical model produces a smooth output pressure (shown in blue) in response to a stepwise desired pressure (shown in orange), closely replicating the dynamics of the pressure applied to the pacifier by the infant—compare the black line in (a) with the blue line in (b).

Main findings of Rochat & Striano

Rochat & Striano [6] tested 0mo and 2mo. Figure 1a illustrates the experimental setup. Each infant was tested in an experiment consisting of 6 phases (for details, see section II-A). The testing started with an initial baseline condition without any contingent auditory stimulation, followed by a sequence of four experimental conditions, alternating analog and non-analog stimulation (in a counterbalanced order). The testing ended with a final baseline condition without any contingent auditory stimulation.

Rochat & Striano [6] found differences in sucking activity of 2-month-old infants compared to newborn infants. First, only 2mo showed differences in sucking activity in the first experimental session (be it analog or non-analog) as compared to the initial baseline. Specifically, compared to 0mo, 2mo' sucking activity had more high-amplitude pressures (more than 0.3 psi) and marginally more just-at-threshold pressures (from 0.1 to 0.125 psi). In the authors' interpretation, only the group of 2mo reacted to the introduction of contingent auditory feedback, whether the sounds were analog or non-analog with respect to the pressure applied on the pacifier. 0mo did not show such a reaction. Second, and most important, only 2mo showed differences in sucking behavior as a function of experimental conditions, analog and non-analog. Specifically, in the analog condition the frequency of sucking activity with pressure levels just at the threshold (from 0.1 to 0.125 psi) was higher than in the non-analog; also, the average pressure amplitude of sucking activity above

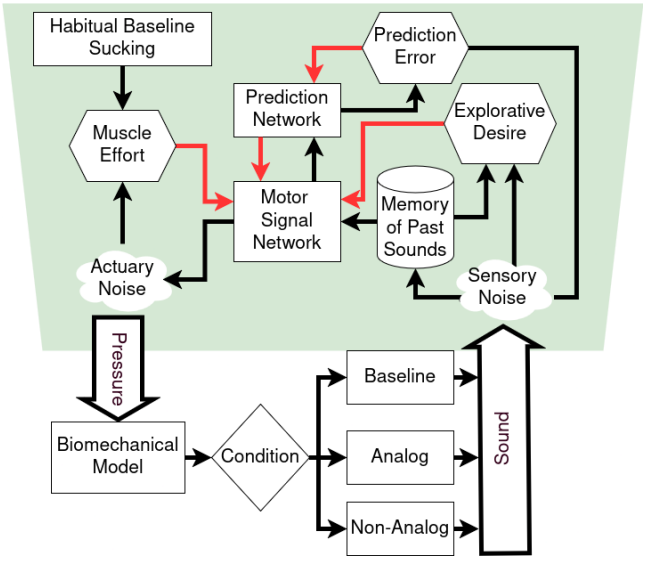


Fig. 2: Overview of the computational model used to study the emergence of self-exploration in 2-month-old infants. The red arrows represent the losses that modify the weights of the neural networks.

threshold (more than 0.1 psi) was lower in the analog than in the non-analog condition. Thus, compared to 0mo, 2mo not only reacted to contingency but were able to modulate their behavior to match the type of contingency, analog or non-analog.

Upon cessation of contingency, some infant studies report a so-called extinction burst – a sudden and transient increase of activity during the final baseline [24], [25]. Rochat & Striano [6] did not report such an extinction burst.

We will evaluate our simulation success by how well they replicate Rochat & Striano [6] findings and other findings, such as the extinction burst.

II. Methods

In this section, we report the technical details of how we modeled infant behavior observed in Rochat & Striano [6]. First, we will describe the experimental setup for our simulations. We then introduce the biomechanical model of the mouth and the sucking behavior. Finally, we describe our two modeling approaches.

A. Experimental setup

To match the number of infants in the original study we ran 20 simulations for both age groups. The original experiments lasted for 540 seconds consisting of six phases of equal length (90 s). We simulated this by having each of our simulations last 5400 steps. With six phases, each lasting 900 steps. As in the original experiment, we implemented two counterbalanced six-phase sequences: 1) baseline $B1$, analog $A1$, non-analog $N1$, analog $A2$, non-analog $N2$, baseline $B2$ and 2) baseline $B1$, non-analog $N1$, analog A1, non-analog $N2$, analog $A2$, baseline $B2$. During the baseline phases, no auditory feedback was generated. From the seven dependent variables used in the original study, we considered those appearing in the main results of the original study: frequency of pressures just at threshold (from 0.1 to 0.125 psi), frequency of high-amplitude pressures (more than 0.3 psi), average of pressures above the 0.1 psi threshold.

B. Model components

Fig. 2 shows the main components of the model introduced in this paper. The simulation starts with receiving an input that represents a sound heard by the baby. Before any learning takes place, the input is modified by the sensory noise, which models a possible inability to accurately perceive sound from the environment. The perceived stimulus is added into the model's memory of past sounds. Note that in our current models, memory refers specifically to the memory of the past sounds and excludes that of past actions and other information; specifically, in our model, memory refers to a value that decides how many past sounds the model uses as input. From the memory of past sounds, the perceived stimulus functions as the input of the motor signal network, a neural network that produces a new motor signal. Motor noise is added to the motor signal to represent potential inaccuracies in the baby's ability to accurately control its sucking motion. Apart from this main path, which produces a new pressure given a perceived sound, the model produces other attributes that are essential for it to learn and act. The novelty preference motivates the model to seek out novel sounds. In contrast, the prediction error motivates accurate predictions which is easier when the actions are known.

Novelty preference utilizes the memory of past sounds and the currently perceived sound to determine the novelty of the current sound. For the prediction error, the second neural network of the model is the most important part. This prediction network has the motor signal from the motor signal network as its input and produces an outcome prediction for the sound that the pacifier will produce based on this motor signal. The prediction is compared with the sound received during the next step to calculate the prediction error. We implemented the preferential level of sucking as a habitual baseline sucking level and computed the muscle effort as the distance between the current sucking and the baseline level. The baseline activity refers to the preferred rhythm of sucking of each infant in the absence of any stimulation. To model it, we used a sine wave and the average of this sine wave is decided by the Strength parameter. Fatigue refers to the negative consequence of going above baseline activity and motivates the model to preserve its strength, thereby stopping it from exerting high pressures constantly.

The sounds produced during the simulations mirror those of the original study. In the baseline condition, no sound is produced. In the analog condition, whenever the exerted pressure exceeds a threshold, the pitch of the produced sound is directly proportional to the pressure exerted on the pacifier. In the non-analog condition, whenever the exerted pressure exceeds a threshold, the pitch of the sound is randomly chosen among a series of sounds.

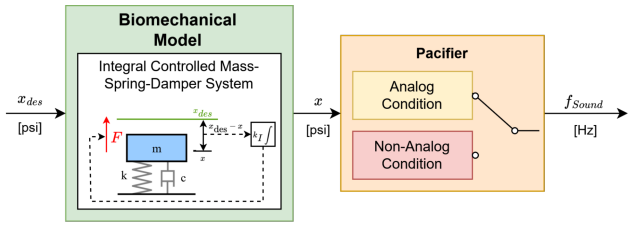


Fig. 3: Overview of the biomechanical model producing a pressure that is applied to the pacifier. The biomechanical model generates a pressure x (measured in psi) by tracking the desired pressure x_{des} generated by the two learning models. This is achieved by simulating the pressure dynamics using a mass-spring-damper system combined with an integral control component. The pacifier produces a sound pitch f_{Sound} analog or non-analog to the applied pressure x.

C. Biomechanical modeling

The dynamic change of pressure inside the mouth is influenced by different factors such as fluid dynamics and biomechanical structure. These dynamics determine how the pressure can change over time, influencing sound generation in the experiment of Rochat & Striano [6], where sounds are generated based on the pressure applied to a pacifier.

To replicate the dynamic behavior from the original study [6] (see Fig. 1a), we developed a simplified biomechanical model based on a mass-spring-damper system combined with an integral control component (see Fig. 3). Mass-spring-damper systems have been used to model biomechanical aspects [26], making them a suitable choice for capturing the intraoral pressure dynamics. This model is used to generate pressures x based on the desired pressures x_{des} output by the learning architectures by simulating the dynamic pressure change observed in [6]. Without a biomechanical model, learning architectures may produce abrupt and stepwise changes in desired pressure x_{des}. By using the biomechanical model, the desired pressures become continuous, and their dynamics resemble that of infants.

The mass-spring-damper system is modeled by the dynamics,

$$a = -d \cdot v + k \cdot (x_{eq} - x) + F_m, \tag{1}$$

where x, v, and a are the position (representing the pressure), velocity, and acceleration of the mass m. The mass m, damping coefficient d, and the spring constant k were empirically selected to replicate the experimental data described in Rochat & Striano [6]. The passive system ($F = 0$) captures the gradual reduction in pressure change rates as the pressure decreases towards the equilibrium pressure x_{eq}. The integral controller $F = k_I \cdot (x_{des} - x)dt$ is used to achieve the gradual increase in pressure change rate during rising pressure. This integral force simulates gradually increasing muscle activation driving the current pressure towards the desired pressure. To reduce oscillations when the pressure reaches the desired pressure level, the integral gain is adapted based on the pressure error $k_I = \tanh(80 \cdot (|x_{des} - x|)) \cdot 0.008$. To integrate sensory and actuator noise in the model, Gaussian noise is added to the desired pressure, the equilibrium pressure, and the forcing term. The resulting behavior of the model is illustrated in Fig. 1b.

D. Learning architectures

We evaluated two different learning architectures for the general model described in Fig. 2: SSL and RL. The key difference between them is how they use external information to adjust their behavior. The SSL model is very controlled and directed towards specific actions while the RL model has less clear supervision and needs to learn from its interaction with the environment.

1) Self-Supervised neural network: We trained a neural network using three motivational factors, novelty preference, prediction error, and effort, mathematically formalized as losses:

1) Novelty preference loss
   $= 1 - \text{MSE}(\text{noisy past sound}, \text{sound})$,
2) Effort loss $= \text{MSE}(\text{noisy force}, \text{instinct force})$,
3) Predictive loss
   $= \text{MSE}(\text{predicted next sound}, \text{true next sound})$.

The novelty preference loss punishes repeated sounds and encourages differences between sounds, the predictive loss encourages correct prediction while the fatigue loss punishes actions diverging from the baseline. The prediction loss is back-propagated through both the prediction model and the motor signal model while the other losses only go through the motor signal model. The mean squared error is used for the effort loss in which the distance between the produced pressure and the pressure that would have been produced by the habitual sucking is measured.

In both networks, the prediction and the motor signal network, consist of fully connected linear layers. The prediction model has one neuron for the input, then ten neurons in the hidden layer and one output neuron. While the number of neurons in each layer is constant in the prediction model, in the motor signal model, the number of neurons depends on memory. The input layer has one neuron per remembered sound, the hidden layer has one more neuron than the input layer and the output layer again has a single neuron. At each step a forward pass is done through these networks to produce a new prediction and action. For the backwards pass the losses are calculated by comparing the output of the model to an optimal output. These losses are propagated backwards through the model allowing the model to improve future outputs. The optimal output for the comparison also stems from the model and evolves over time.

2) Reinforcement learning: In RL an agent interacts with an environment and receives a reward for each action. Its goal is to maximize the cumulative reward. The problem is mathematically formalized as a Markov Decision Process (MDP) which is defined by a tuple $(\mathcal{S}, \mathcal{A}, P, r, \rho_0, \gamma)$, where $\mathcal{S}$ is the set of states, $\mathcal{A}$ is the set of actions. $P(s'|s, a)$ is the probability of transitioning from state s into state s' given the action a. $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function, ρ_0 is the initial state distribution and $\gamma \in (0, 1)$ is the discount factor,

representing the difference in importance between future and present rewards.

Furthermore, to capture the different phases of Rochat & Striano's [6] experiment (baseline, non-analog, analog) we extend the MDP by a contextual formalism (Contextual Markov Decision Process, CMDP [27]). A CMDP is defined by a tuple $(\mathcal{C}, \mathcal{S}, \mathcal{A}, m)$, where $\mathcal{C}$ is the context space and m is a function that maps a context $c \in \mathcal{C}$ to an MDP $m(c) = (\mathcal{S}, \mathcal{A}, P^c, r^c, \rho_0^c, \gamma)$. A CMDP thus defines a family of MDPs, that all share an action and state space, but the transition probability P^c, the reward function r^c and the initial state distribution ρ_0^c may differ depending on the context c. In our case just P^c differs w.r.t. the context which is the experimental phase.

The RL agent can apply pressure (action) to the pacifier. The actions follow the biomechanical model of a Mass-Spring-Damper system (Fig. 3). The environment transforms the pressure to the corresponding sound frequency and returns it as a state to the agent. As the output action (pressure) of the agent and the internal state (sound frequency) are continuous, we used the off-policy algorithm Soft-Actor Critic (SAC) [28] for our RL optimization.

The reward formulation is analogous to the three loss terms of the SSL architecture with an opposite sign. Reward components 1 and 2 are covered through the environment. Reward 3 is intrinsic to the agent, and is considered separately with a forward model $\tilde{s}_{t+1} = f(s_t, a_t)$ which is trained in parallel to the RL updates [29].

Since the RL model learns through interactions with the environment and receives less direct supervision from rewards, it is expected to have slower learning as compared to SSL. We confirmed this in an ablation trial where both models underwent the same number of learning steps. To account for the slower learning, we conducted ten iterations of the six experimental phases per infant simulation.

E. Parameter tuning

In both architectures, we have seven hyperparameters (or developmental factors) that govern learning: i) prediction of the outcomes, ii) novelty preference, iii) fatigue, iv) strength, v) memory, vi) sensory noise, and vii) motor noise. Rather than arbitrarily choosing values to represent the age groups, we used a grid search to go through all combinations of them, analyzing which parameters lead to a high frequency of pressure around the threshold. Results show that memory is the most important developmental factor to differentiate the two age groups. To implement this, the memory of the simulated 0mo was set to include only the last sound while that of the 2mo included the last five sounds.

To determine which developmental factor is most important in each model architecture, we computed the Pearson correlation coefficients between the seven developmental factors and the main outcome of [6], that is, the difference of pressure threshold hits between the analog and non-analog phases. We do not consider strength and fatigue for this computation (by setting fatigue to zero) as both interact with the desired pressure threshold value without contributing to learning. For successful simulation, the pressure needs to approach the threshold more frequently in the analog condition than in the non-analog one. However, the strength parameter defines the starting pressure level and hence it determines whether the pressure needs to increase or decrease from the initial level in order to reach the threshold. Therefore, strength impacts the activity of the model but does not necessarily affect learning. Fatigue determines how fast the simulation can approach the threshold and it also mostly affects model activity but not necessarily learning.

In the SSL model, the factors with the highest correlation coefficients are motor noise (0.28) and memory (0.18). In the RL model, the factors with the highest correlation coefficients are novelty preference (-0.66), memory (0.28), and sensory noise (0.17).

In both our models, noise mechanically hinders the models from approaching the desired pressure threshold. Noise prevents the simulated agent from control and activity modulation. We therefore kept it constant for both age groups. Also, note that in the RL model there is an inbuilt novelty preference (see the entropy-based exploration term in SAC [28]).

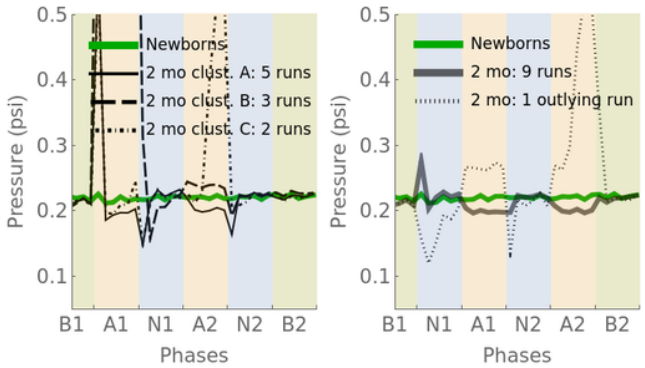
In conclusion, memory appears to be the common developmental factor that affects the learning in both models. Based on this finding, we decided to differentiate the age groups based on their memory, with 2mo remembering the last five sounds while 0mo only remember the last sound. Additionally, the 2mo have a novelty preference of 1 while 0mo have it at 0. Novelty preference in the SSL model increases the overall amount of threshold hits, although it does not have a high correlation to more analog than non-analog threshold hits. In the RL model we observe a negative correlation between the novelty preference and the amount of threshold hits; it may be due to our novelty preference term interfering with the inbuilt novelty term in the RL model.

III. Results

Before any specific comparisons and choosing a specific dependent variable (among the 7 suggested in the original study) to characterize an aspect of activity, we will first investigate whether the overall activity is modulated across different experimental phases and differently so for each age group. We will then describe how well our models replicate the comparisons between conditions that Rochat & Striano [6] reported as significant.

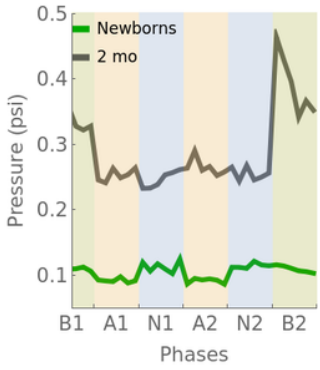
A. Modulation of activity – replication of general finding in Rochat & Striano

Fig. 4 shows the average activity (across individual simulation runs) of 0mo and 2mo, separately for conditions with analog and non-analog start. We show both the results of the SSL model and the RL model. Both models show clear differences between the age groups regardless of the starting phase. Not only do we observe a difference in age groups but we also clear differences between the phases in the 2mo. In the RL model, at the start of the 2nd baseline (B2), we clearly observe an extinction burst for the 2mo, which is not observed for the SSL model. We further find that the

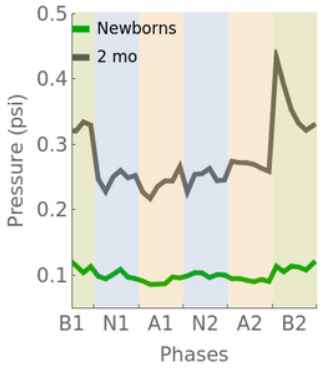


(a) SSL model Analog start

(b) SSL model Non-Analog start



(c) RL model Analog start



(d) RL model Non-Analog start

Fig. 4: Activity per age group, separately for infants starting with an analog or a non-analog phase after the 1st baseline, separately for SSL and RL models. The activity was binned in 150-step bins, corresponding to 15-second bins as in Rochat & Striano [6]. The lines show the means across individual runs, green for 0mo and black or gray for 2mo (2 mo). Experimental phases are highlighted with color backgrounds. $B1$: 1st baseline; $A1$ and $A2$: 1st and 2nd analog experimental phases; $N1$ and $N2$: 1st and 2nd non-analog experimental phases; $B2$: 2nd baseline. Note that only the second half of $B1$ is shown. Results of (a) SSL simulations starting with $A1$. The activity of the 2mo reveals 3 clusters, shown with separate black lines. (b) SSL simulations starting with $N1$. The atypical activity of 1 simulation run is shown with a thin dotted line. The thick gray line shows the average activity of 9 simulation runs. (c) RL simulations starting with $A1$. (d) RL simulations starting with $N1$.

activity of the SSL simulated 2mo can be clustered into three distinct groups shown in Fig. 4. These clusters mostly appear in runs that start with the analog phase, we hypothesize that the reason for this is that there is more to learn from the analog phases leading to different local optima being found by the model rather than finding a more general solution as observed in non-analog starts.

Tab. I shows statistics of different comparisons reported in this section. For the SSL model, 0mo show no difference in activity between the first baseline ($B1$) and the first experimental phase ($E1$: either analog $A1$ or non-analog $N1$; comparison $E1 \neq B1$) while the activity of 2mo is clearly different in the two phases. For the RL model, both 0mo and 2mo show differences in activity between the first baseline and the subsequent first experimental phase. Thus, for 2mo, both models show a difference of activity between the baseline and the first experimental phase. For 0mo, both models show no or a smaller difference.

In the SSL model, as expected, 0mo don't show any difference in activity between the two analog phases ($A1$ and $A2$) and the two non-analog ones ($N1$ and $N2$; comparison $A \neq N$), while the 2mo show clear differences. The direction of the modulation of activity in 2mo depends on the starting experimental condition. For the RL model, both 0mo and 2mo show differences in activity between the two analog and the two non-analog phases with the direction of modulation, similar to the SSL model, depending on the starting condition. Thus, as in the previous comparison ($E1 \neq B1$), for 2mo both types of models show a difference between the activity during the two analog and the two non-analog phases but for 0mo only the RL model shows a difference.

For the SSL model, when we compare the Post-extinction and Pre-extinction phases, only a small number of 0mo show any difference. However, for the 2mo, we observe a difference when the Pre-extinction phase is analog. For the RL model, an extinction burst is clearly present in most 2mo; in 0mo the amplitude of the extinction burst is lower by an order of magnitude. Thus, for 2mo both models show extinction bursts but for 0mo they both show no or greatly decreased extinction bursts.

B. Replication of specific results in Rochat & Striano

The three main results in Rochat & Striano [6], focus on 1) comparisons between the frequency of high-amplitude pressures in the phase $E1$ and $B1$, 2) the frequency of just-at-threshold pressures in the conditions A and N, and 3) the average pressure amplitude above threshold in the conditions A and N. We show the results of our model for these three comparisons in Tab. II.

We will first verify whether the frequency of high-amplitude pressures is higher in the first experimental session than in the first baseline ($E1 > B1$). For the SSL model, the average frequency is higher in $E1$ than in $B1$, and, as expected, to a much larger extent for the 2-month-old infants. For the RL model, the result is very similar. Thus, both models replicate this finding of Rochat & Striano [6] well.

Next, we will look whether there are more just-at-threshold pressures in the analog than in the non-analog phases. For the SSL model, that is the case on average for both 0mo and 2mo but the difference is one order of magnitude larger for 2mo compared to 0mo and there are more 2mo for whom this difference is statistically significant. For the RL model, the conclusion is the same, though the differences between the two age groups are smaller. Thus, the second finding of Rochat & Striano [6] is also, on a descriptive level, replicated by both models.

Comparison	Model	0mo	2mo
1st experimental phase vs 1st baseline $E1 \neq B1$ $(E1 - B1)$	SSL	.0002 ± .003 n = 20 (*2)	**.139 ± .28 n = 20 (*18)**
	RL	-.015 ± .015 n = 20 (*16)	**.077 ± .07 n = 20 (*19)**
$E1 > B1$	SSL	.002 ± .001 n = 14 (*1)	.168 ± .295 n = 17 (*16)
	RL	.007 ± .005 n = 4 (*2)	.04 ± .03 n = 2 (*2)
$E1 < B1$	SSL	-.003 ± .001 n = 6 (*1)	-.024 ± .019 n = 3 (*2)
	RL	-.02 ± .012 n = 16 (*14)	-.09 ± .06 n = 18 (*17)
Analog vs Non-Analog $A \neq N$ $(N - A)$	SSL	.001 ± .002 n = 20 (*5)	**-.07 ± .15 n = 20 (*20)**
	RL	.015 ± .016 n = 20 (*18)	**-.007 ± .036 n = 20 (*16)**
$A < N$	SSL	.002 ± .001 n = 14 (*3)	.024 ± .013 n = 10 (*10)
	RL	.02 ± .015 n = 16 (*15)	.023 ± .022 n = 9 (*6)
$A > N$	SSL	-.001 ± .001 n = 6 (*2)	-.164 ± .166 n = 10 (*10)
	RL	-.004 ± .002 n = 4 (*3)	-.032 ± .026 n = 11 (*10)
Post-Extinction > Pre-Extinction	SSL	.004 ± .003 n = 13 (*5)	**.015 ± .01 n = 12 (*10)**
	RL	.021 ± .015 n = 16 (*15)	**.236 ± .102 n = 17 (*17)**

TABLE I: Modulation of activity across conditions and age groups. Each cell includes the average ± standard deviation of the individual differences of pressure (in psi units) between conditions for each specific comparison; next, the number of differences considered is given and the (*) shows how many of those individual differences were statistically significant (bootstrap with 10^5 resamples). For Pre- vs. Post-Extinction, we included 20% of data before and after the extinction at the start of the 2nd baseline $B2$.

Finally, we will look at whether the average pressure amplitude above threshold is lower in the analog than in the non-analog phases. For the SSL model, that is the case on average for both 0mo and 2mo but the difference is larger for 2mo compared to the 0mo. For the RL model, while the results are in the expected direction, the difference is larger for 0mo compared to 2mo. Thus, the third finding of Rochat & Striano [6] is partially replicated by both models.

C. Comparison between the two learning architecture implementations

With some exceptions, the models "behave" similarly. Except for the first comparison ($E1 \neq B1$), in both models, the activity increases or decreases across conditions in a similar fashion. However, there are a few differences between them. On the one hand, the SSL model differentiates the two age groups better. On the other hand, the RL model generates a larger and more reliable extinction burst. Infant learning likely includes more than one learning mechanism. In future work, combining the two models could allow for a closer simulation of infant behavior.

Comparison	Model	0mo	2mo
$E1 > B1$ Frequency of high-amplitude pressures	SSL	**4 ± 4 n = 6 (*2)**	**171 ± 315 n = 20 (*20)**
	RL	**5 ± 2 n = 2 (*1)**	**76 ± 77 n = 6 (*6)**
$A > N$ Frequency of just-at-threshold pressures	SSL	**34 ± 49 n = 13 (*2)**	**306 ± 237 n = 13 (*12)**
	RL	**32 ± 40 n = 9 (*4)**	**51 ± 39 n = 17 (*9)**
$A < N$ Average pressure amplitude above threshold (psi)	SSL	**.002 ± .002 n = 15 (*8)**	**.023 ± .012 n = 12 (*11)**
	RL	**.037 ± .022 n = 20 (*18)**	.022 ± .021 n = 9 (*7)

TABLE II: Target comparisons with Rochat & Striano [6]. Also see description of Tab. I.

IV. Discussion

We successfully modeled the infants' behavioral pattern described by Rochat & Striano [6], using both SSL and RL architectures. Both architectures are embodied with respect to the mouth due to the biomechanical module that implements real-world physical constraints of oral pressure dynamics. This results in dynamic motor actions and consequently sensory stimulation being more consistent over time. Results of both architectures converged on increased memory capacity of action effects as the critical developmental factor underlying the active exploration behavior and thus agency. We analyzed how the overall activity is modulated across phases and if this modulation differs across age groups. We found that, compared to 0mo, the 2mo show both larger modulations of activity and the modulations are statistically significant for a higher proportion of individuals. Additionally, the direction of 2mo' modulations varies depending on which of the two sequences they were subjected to. Furthermore, when focusing on the three specific and statistically significant differences reported by Rochat & Striano [6], both learning architectures replicated all of them. Finally, the difference between the two age groups seems to be better replicated by the SSL than by the RL one. Rochat & Striano [6] emphasized the difference of activity between the analog and the non-analog phases and our models replicated both the general activity modulation and that of specific dependent variables selected by Rochat & Striano [6]. Interestingly, we also observed spontaneously emerging extinction bursts in both architectures.

Both architectures implemented the same higher-level developmental factors of the overall model (see Fig. 2). They include i) outcome prediction, ii) novelty preference, iii) fatigue, iv) strength, v) memory, vi) sensory noise, and vii) motor noise. Crucially, despite the models being based on different algorithmic learning mechanisms, SSL or RL, both models end up with the same general conclusion: increasing memory of action effects is the developmental factor that best explains the difference between the behavior of our simulated 0mo and 2mo. For 0mo we set the memory to remember only the last sound while 2mo were able to remember five sounds. Without such action-effect memory, identifying the nature of longer sequences, necessary to distinguish analog from non-

analog sequences, would be impossible. Distinguishing the phases allows the model to adapt to them, therefore memory is the key developmental factor in both models. Another developmental factor of interest was novelty preference. We observed that it was only needed in the SSL model. Compared to the SSL model, in the RL model the novelty preference had an opposite effect. One potential explanation for this is that in RL architectures an exploration term is already embedded and our own novelty preference may interfere with it. Here we have given an initial idea of the developmental factors at play in replicating the findings of Rochat & Striano [6] and how they influence learning. We have not considered strength and fatigue during parameter tuning despite them playing a role in the models' results because of their direct connection to the baseline and threshold values. However, further ablation studies, which focus on one parameter at a time, will allow a more exact and in-depth view on how each parameter affects learning sensorimotor contingencies.

In summary, using two different architectures (SSL & RL) that implemented the same simplified overall model of infant sensorimotor control, we showed that the results of Rochat & Striano [6] can be replicated. Based on plausible developmental factors, both architectures (SSL & RL) were able to differentiate between experimental conditions and adapt activity in ways that closely resemble infant behavior. Although not a new insight per se (e.g. [15]), we generated converging evidence that action-effect memory plays a crucial role for self-exploration. Indeed, other ideomotor theorists such as [30] have suggested that, besides representing the current situation, a secondary representation of the desired goal (that may come from memory) is crucial for agency and voluntary action. In conclusion, the presented computational perspective on the emergence of active exploration suggests a renewed emphasis on action-effect memory for early sensorimotor contingency learning and developing action control.

REFERENCES

[1] W. James, F. Burkhardt, F. Bowers, and I. K. Skrupskelis, *The principles of psychology*. Macmillan London, 1890, vol. 1, no. 2.

[2] S. Gallagher, "Philosophical conceptions of the self: implications for cognitive science," *Trends in cognitive sciences*, vol. 4, no. 1, pp. 14–21, 2000.

[3] S. A. Verschoor and B. Hommel, "Self-by-doing: The role of action for self-acquisition," *Social Cognition*, vol. 35, no. 2, pp. 127–145, 2017.

[4] M. Liesner, N.-A. Hinz, and W. Kunde, "How action shapes body ownership momentarily and throughout the lifespan," *Frontiers in Human Neuroscience*, vol. 15, p. 697810, 2021.

[5] N.-A. Kollakowski, M. Mammen, and M. Paulus, "What is the implicit self in infancy? a classification and evaluation of current theories on the early self," *Cognitive Development*, vol. 68, p. 101394, 2023.

[6] P. Rochat and T. Striano, "Emerging self-exploration by 2-month-old infants," *developmental science*, vol. 2, no. 2, pp. 206–218, 1999.

[7] S. A. Verschoor, M. Spapé, S. Biro, and B. Hommel, "From outcome prediction to action selection: developmental change in the role of action–effect bindings," *Developmental Science*, vol. 16, no. 6, pp. 801–814, 2013.

[8] S. Verschoor, M. Weidema, S. Biro, and B. Hommel, "Where do action goals come from? evidence for spontaneous action–effect binding in infants," *Frontiers in Psychology*, vol. 1, p. 201, 2010.

[9] L. Zaadnoordijk, M. Otworowska, J. Kwisthout, S. Hunnius, and I. van Rooij, "The mobile-paradigm as measure of infants' sense of agency? insights from babybot simulations," in *2016 Joint IEEE international conference on development and learning and epigenetic robotics (ICDL-EpiRob)*. IEEE, 2016, pp. 41–42.

[10] S. Verschoor and S. Biro, "Primacy of information about means selection over outcome selection in goal attribution by infants," *Cognitive science*, vol. 36, no. 4, pp. 714–725, 2012.

[11] S. Biro, S. Verschoor, E. Coalter, and A. M. Leslie, "Outcome producing potential influences twelve-month-olds' interpretation of a novel action as goal-directed," *Infant Behavior and Development*, vol. 37, no. 4, pp. 729–738, 2014.

[12] E. Mather, "Novelty, attention, and challenges for developmental psychology," *Frontiers in psychology*, vol. 4, p. 491, 2013.

[13] S. Adams, D. Jones, A. Esmail, and E. Mitchell, "What affects the age of first sleeping through the night?" *Journal of paediatrics and child health*, vol. 40, no. 3, pp. 96–101, 2004.

[14] T. A. Davis and M. L. Fiorotto, "Regulation of muscle growth in neonates," *Current Opinion in Clinical Nutrition & Metabolic Care*, vol. 12, no. 1, pp. 78–85, 2009.

[15] C. Rovee-Collier and K. Cuevas, "The development of infant memory," in *The development of memory in infancy and childhood*. Psychology Press, 2008, pp. 23–54.

[16] A. J. DeCasper and W. P. Fifer, "Of human bonding: Newborns prefer their mothers' voices," *Science*, vol. 208, no. 4448, pp. 1174–1176, 1980.

[17] J. Mehler, P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini, and C. Amiel-Tison, "A precursor of language acquisition in young infants," *Cognition*, vol. 29, no. 2, pp. 143–178, 1988.

[18] C. Moon, R. P. Cooper, and W. P. Fifer, "Two-day-olds prefer their native language," *Infant behavior and development*, vol. 16, no. 4, pp. 495–500, 1993.

[19] M. Hadders-Algra, "Early human motor development: From variation to the ability to vary and adapt," *Neuroscience & Biobehavioral Reviews*, vol. 90, pp. 411–427, 2018.

[20] N. J. Butko and J. R. Movellan, "Detecting contingencies: An infomax approach," *Neural Networks*, vol. 23, no. 8-9, pp. 973–984, 2010.

[21] 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, pp. 389–421, Jul. 2014.

[22] F. Mannella, V. G. Santucci, E. Somogyi, L. Jacquey, K. J. O'Regan, and G. Baldassarre, "Know Your Body Through Intrinsic Goals," *Frontiers in Neurorobotics*, vol. 12, p. 30, 2018.

[23] P. Lanillos and G. Cheng, "Adaptive robot body learning and estimation through predictive coding," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4083–4090.

[24] S. M. Alessandri, M. W. Sullivan, and M. Lewis, "Violation of expectancy and frustration in early infancy," *Developmental Psychology*, vol. 26, no. 5, pp. 738–744, 1990, place: US Publisher: American Psychological Association.

[25] J. C. Heathcock, A. N. Bhat, M. A. Lobo, and J. C. Galloway, "The Relative Kicking Frequency of Infants Born Full-term and Preterm During Learning and Short-term and Long-term Memory Periods of the Mobile Paradigm," *Physical Therapy*, vol. 85, pp. 8–18, 2005.

[26] P. E. Hammer, M. S. Sacks, P. J. del Nido, and R. D. Howe, "Mass-Spring Model for Simulation of Heart Valve Tissue Mechanical Behavior," *Annals of biomedical engineering*, vol. 39, no. 6, pp. 1668–1679, Jun. 2011.

[27] A. Hallak, D. Di Castro, and S. Mannor, "Contextual markov decision processes," *arXiv preprint arXiv:1502.02259*, 2015.

[28] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. Pmlr, 2018, pp. 1861–1870.

[29] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-driven exploration by self-supervised prediction," in *International conference on machine learning*. PMLR, 2017, pp. 2778–2787.

[30] W. Prinz, *Open minds: The social making of agency and intentionality*. MIT Press, 2012.