Curiosity-Driven Development of Tool Use Precursors: a Computational Model

Sébastien Forestier, Pierre-Yves Oudeyer

Curiosity-Driven Development of Tool Use Precursors: a Computational Model

Sébastien Forestier (sebastien.forestier@inria.fr)
Inria Bordeaux Sud-Ouest and Université de Bordeaux, France

Pierre-Yves Oudeyer (pierre-yves.oudeyer@inria.fr)
Inria Bordeaux Sud-Ouest and Ensta-ParisTech, France

Abstract

Studies of child development of tool use precursors show successive but overlapping phases of qualitatively different types of behaviours. We hypothesize that two mechanisms in particular play a role in the structuring of these phases: the intrinsic motivation to explore and the representation used to encode sensorimotor experience. Previous models showed how curiosity-driven learning mechanisms could allow the emergence of developmental trajectories. We build upon those models and present the HACOB (Hierarchical Active Curiosity-driven mOdel Babbling) architecture that actively chooses which sensorimotor model to train in a hierarchy of models representing the environmental structure. We study this architecture using a simulated robotic arm interacting with objects in a 2D environment. We show that overlapping phases of behaviours are autonomously emerging in hierarchical models using active model babbling. To our knowledge, this is the first model of curiosity-driven development of simple tool use and of the self-organization of overlapping phases of behaviours. In particular, our model explains why and how intrinsically motivated exploration of non-optimal methods to solve certain sensorimotor problems can be useful to discover how to solve other sensorimotor problems, in accordance with Siegler's overlapping waves theory, by scaffolding the learning of increasingly complex affordances in the environment.

Keywords: curiosity-driven learning; tool use; goal babbling; overlapping waves; developmental trajectory; HACOB model

Introduction

The understanding of tool use development in young children is one of the key question for the more general understanding of the ontogeny of human cognition. Indeed, a series of abilities are progressively developed from the simplest reaching movements of the arms through more dexterous manipulation of a spoon, towards advanced control of multiple interacting objects. The latter shows an understanding of shapes, forces and other physical properties that can be hierarchically recruited for mental transformations and planning operations which are pillars of human cognition. Child development has first been described as staircase-like successive stages in which all children go through (Piaget, Cook, & Norton, 1952). More recently, different views were developed and describe the structure and variability of observed children's developmental paths. In particular, the development of tool use precursors can be described as three consecutive and overlapping stages of behaviours where sequential learning and goal-directed behaviours play an increasing role (Guerin, Kruger, & Kraft, 2013): body babbling, behaviours with a single object, and behaviours with several interacting objects. A study of free play (Zelazo & Kearsley, 1980) shows that at $9\frac{1}{2}$ months play is mostly composed of tactile examination, waving or mouthing of a single object but simple relational acts of banging two objects together are already present. Later at $13\frac{1}{2}$ months, the study reveals that most children instead prefer to explore the relationships among objects, but still show behaviors of the previous phase. Furthermore, they show that this overlapping phases pattern averaged across children is also present in a longitudinal study of a single child. Another type of behavioral structure is described by Siegler as the variability in a child's set of current methods to solve a problem, which leads to the overlapping waves theory (Siegler, 1996).

In this paper we focus on the study of such progressions between phases of behaviours in a robotic model, and on the use of concurrent methods to solve a problem. We hypothesize that several mechanisms play a role in behavioural progression's structure and in particular 1) the intrinsic motivation to explore through a self-regulation of the growth of complexity of self-selected skills or tasks; 2) the structure of the representation used to encode sensorimotor experience.

Intrinsic motivation, sometimes called "curiosity", have been identified to play a fundamental role in driving spontaneous exploration in infant free play (Kidd & Hayden, 2015). They have been defined as mechanisms that push infants to explore activities for their own sake, driven by the search of novelty, surprise, dissonances or optimal challenge (Gottlieb, Oudeyer, Lopes, & Baranes, 2013). In the last decade, various families of computational models of intrinsic motivation were developed, often based on the formal frameworks of active learning and reinforcement learning (Baldassarre & Mirolli, 2013). One family of models, that has targeted to study the developmental dimensions of intrinsic motivation, has considered a curiosity-driven learning mechanism where the learner actively engages in sensorimotor activities that provide high learning progress, avoiding situations that are too easy or too difficult and progressively focusing on activities of increasing complexity (Gottlieb et al., 2013). Such computational models have shown that developmental trajectories could emerge from the curiosity-driven learning of sensorimotor mappings, in very different settings. In the Playground Experiment (Oudeyer, Kaplan, & Hafner, 2007), a quadruped robot motivated to maximize its learning progress acquired how to use its motor primitives to interact with the items of an infant play mat and a robot peer, following a self-organized learning curriculum. In (Baranes & Oudeyer, 2013), such mechanisms were shown to allow for efficient learning of large repertoires of skills involving high-dimensional continuous actions, as intrinsic motivation guided the system to explore sensorimotor problems of increasing complexity. In a model of active vocal development (Moulin-Frier, Nguyen, & Oudeyer, 2014), an agent had to learn how to produce sounds with its vocal tract by self-exploration combined with imitation of adult speech sounds. This model reproduces accurately major phases of infant vocal development until 6 months. In both studies, developmental trajectories are emerging from learning, with both regularities in developmental steps and diversity. Regularities result from the attractor dynamics of the interaction between motivated learning, the body and the environment. The diversity comes from different mechanisms: stochasticity in the algorithms, variability in the environment, and the multiple attractors of the dynamic learning system. Existing models have considered the exploration and learning of sensorimotor correspondences mapping a motor space to a single task/sensory space. However, in the perspective of an open-ended development of reusable skills, and specifically in the development of tool use, multiple interdependent and hierarchically organized task spaces should be available to the agent. For instance, using a tool to act upon an object could make use of previously explored interaction with the tool. Consequently, an intrinsic motivation towards learning progress maximization could particularly be useful in the context of tool use where progress on some high-level task can not happen before progress on lower-level tasks have been made, by focusing training on currently learnable self-generated tasks.

We study aspects of those hypothesis leveraging previous models of curiosity-driven learning and extending them to the active exploration of hierarchical sensorimotor and task spaces. We define a hierarchy of sensorimotor models that structures the sensory space to reflect the interaction of the different items of the environment. The question of the autonomous learning of such a sensorimotor hierarchy is an important one but is not essential to test the questions asked in this paper, so here we provide the hierarchy to the agent as a prior. In this hierarchy of models to explore, different exploration choices are available to the agent: which model to explore, and how to explore that model. The problem of finding an efficient active choice strategy is an instance of strategic learning (Nguyen & Oudeyer, 2012), where different outcomes and strategies are available and the agent has to learn which strategies are useful for which outcomes. This can be viewed as a generalization of active learning methods in machine learning. We define the HACOB (Hierarchical Active Curiosity-driven mOdel Babbling) architecture and compare several possible strategies to study the role of active learning and hierarchical representation in the structuring of developmental trajectories. We compare the different learning conditions in a 2D environment where a simulated arm with three joints plus a gripper can grab one of two available tools to move an out-of-reach object, and we study the structure of behavioural phases during exploration.

To our knowledge, HACOB is the first model of the curiosity-driven development of tool use, and the first to show the autonomous emergence of overlapping phases in the development of simple tool use in a simulated robotic setup. Here we define tool use as the ability to perform different effects on an object with the help of an intermediate object, using some sort of learned inverse mapping. Our model is also the first to account for the intrinsically-motivated parallel exploration of different tools to reach one goal, in line with Siegler's overlapping waves theory. Other models predefine successive phases in object affordances learning (Ugur, Nagai, Sahin, & Oztop, 2015), or do not study the role of intrinsic motivation in tool affordances learning (Stoytchev, 2005), or have only considered the autonomous development of single object manipulation (Gottlieb et al., 2013).

However, here we do not study some important factors in the development of tool use. For instance, young infants need to adapt to the maturation of vision and to a developing body. Also, social guidance through imitation and mimicry is of central importance for the development of tool use but we do not address the question of its modeling in this paper nor of the interplay between social learning and self-exploration.

Methods

Environment

We simulate[1] a 2D robotic arm that can grasp tools that can be used to move an object into different boxes in the environment. In each trial, the agent executes a motor trajectory and gets the associated sensory feedback. Finally the arm, tools and objects are resetted to their initial state. The next sections precisely describe the items of the environment and their interactions. See Fig.1 for an example state of the environment.

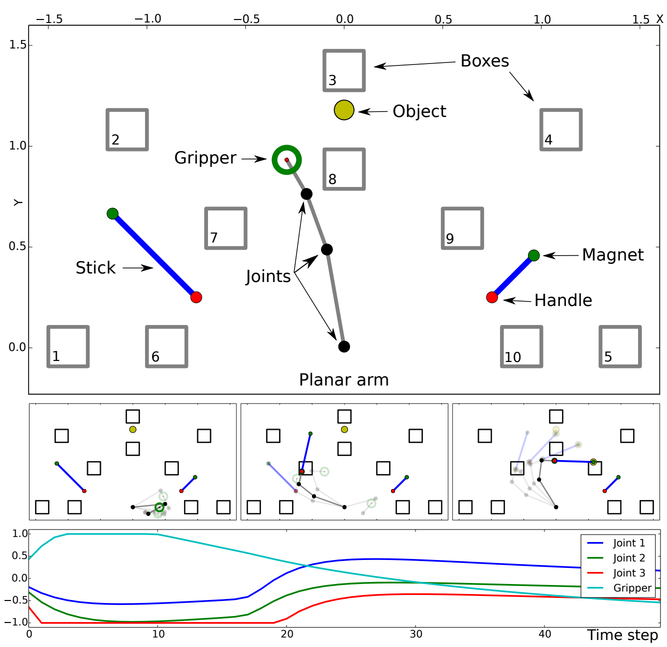


Figure 1: Top: a state of the environment. Middle: position of the arm at time steps 17, 33 and 50, with some intermediate positions, along the 50 steps movement. Bottom: trajectory of each of the four virtual motors, generated by a DMP.

[1]Source code and notebooks available as a Github repository at https://github.com/sebastien-forestier/CogSci2016

Robotic Arm The 2D robotic arm has 3 joints plus a gripper located at the end-effector. Each joint can rotate from $-\pi$ *rad* to π *rad* around its resting position, mapped to a standard interval of $[-1,1]$. The length of the 3 segments of the arm are 0.5, 0.3 and 0.2 so the length of the arm is 1 unit. The resting position of the arm is vertical with joints at 0 *rad* and its base is fixed at position $[0,0]$. The gripper g has 2 possible positions: *open* ($g \geq 0$) and *closed* ($g < 0$) and its resting position is *open* (with $g = 0$). The robotic arm has 4 degrees of freedom represented by a vector in $[-1,1]^4$. A trajectory of the arm will be represented as a sequence of such vectors.

Motor Control We use Dynamical Movement Primitives (Ijspeert, Nakanishi, Hoffmann, Pastor, & Schaal, 2013) to control the arm's movement as this framework permits the production of a diversity of arm's trajectories with few parameters. Each of the 4 arm's degrees-of-freedom (DOF) is controlled by a DMP starting at the rest position of the joint. Each DMP is parameterized by one weight on each of 2 basis functions and one weight specifying the end position of the movement. The weights are bounded in the interval $[-1,1]$ and allow each joint to fairly cover the interval $[-1,1]$ during the movement. Each DMP outputs a series of 50 positions that represents a sampling of the trajectory of one joint during the movement. The arm's movement is thus parameterized with 12 weights, represented by the motor space $M = [-1,1]^{12}$.

Objects and Tools Two sticks can be grasped by the handle side in order to catch an out-of-reach object. A small stick of length 0.3 is located at position $(0.75, 0.25)$ and a long stick of length 0.6 is located at position $(-0.75, 0.25)$ as in Fig. 1. An object (yellow ball), initially at position $(0, 1.2)$, can be caught by the magnetic side of one of the two sticks, moved and possibly placed into one of ten fixed squared boxes. If the gripper is closed near the handle of a stick (closer than 0.2), it is considered grasped and follows the gripper's position and the angle of the arm's last segment until the gripper opens. Similarly, if the magnetic side of a stick reaches the ball (within 0.1), the ball will then follow the magnet. The ten boxes (identified from 1 to 10) are static and have size 0.2. Boxes 1 to 5 can only be reached with the long stick, and the other five boxes can be reached with both sticks.

Sensory Feedback At the end of the movement, the robot gets sensory feedback from the different items of the environment (S, 25D). First, the trajectory of the gripper is represented as the x and y positions and the aperture (1 or -1) of the gripper at 3 time points: steps 17, 33, 50 during the movement of 50 steps (S_{Hand}, 9D). Similarly, the trajectories of the end points of the sticks are 3-point sequences of x and y positions (S_{Stick_1} and S_{Stick_2}, 6D each). It also gets the position of the single object at the end of the movement (S_{Object}, 2D). The agent receives the identifier (from 1 to 10) of the reached box if one of them has been reached by the ball, 0 otherwise. It also receives the distance between the ball at the end of the movement and the closest box (S_{Boxes}, 2D).

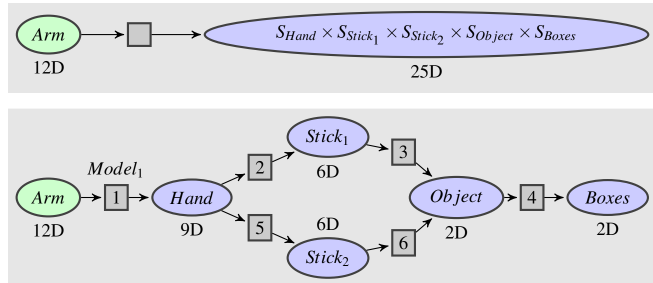


Figure 2: Architectures. Top: Flat. Bottom: Hierarchical.

Learning Architectures

The problem settings for the learning agent is to explore its sensorimotor space and collect data so as to discover how to produce a diversity of effects, and to learn repertoires of skills allowing to reproduce these effects in the form of inverse models. Consequently, the system is not given a priori a single target task to be solved: it rather autonomously selects the sensorimotor problems it will focus on through an intrinsically motivated selection of sensorimotor models.

Flat Architectures We define a flat architecture as directly mapping the motor space M (12D) and the sensory space S (25D). To do so, the agent needs a sensorimotor model that learns the mapping and provides inverse inference of a probable m to reach a given s. The sensorimotor model stores new information of the form (m,s) with $m \in M$ being the experimented motor parameters and $s \in S$ the associated sensory feedback. It computes the inverse inference with the nearest neighbour algorithm: it gets the motor part of the nearest neighbour in S of the given s, and adds exploration noise (Gaussian with $\sigma = 0.01$) to explore new motor parameters.

The agent also needs an interest model that chooses goals in the sensory space. The control condition is a random motor babbling condition (F-RmB) that always randomly chooses new motor parameters m. In the other conditions, the agent performs Goal Babbling, a method by which it self-generates goals in the sensory space and tries to reach them. To generate those goals, different strategies have been studied (Baranes & Oudeyer, 2013). It was shown that estimating the learning progress in different regions of the sensory space and generating the goals where the progress is high leads to a fast learning. However, this cannot be applied in a 25D sensory space as a learning progress signal cannot be estimated in this volume. Thus, in the flat random goal babbling condition (F-RGB), we use a random generation of goals in the sensory space, which was nevertheless proven to be highly efficient in complex sensorimotor spaces (Rolf, Steil, & Gienger, 2010).

Hierarchical Architectures The 25D sensory space can be structured to reflect the interaction of the different items of the environment. Indeed, the arm motor position influences the gripper, which influences one of the tools (but not both at the same time), which influences the position of the object and the filling of the boxes. We thus study here learning architectures that could make use of this sensorimotor structure, and we call them hierarchical. Those architecture learn 6 models at the same time (see Fig. 2: gray squares are models). Each of those models functions in the same way as the random goal babbling flat architecture (F-RGB). Each model has a motor space (e.g. motor space of model 2 is S_{Hand}), a sensory space (respectively S_{Stick_1}, see arrows in Fig. 2), and can choose goals randomly in this sensory space. At each iteration, the architecture first has to choose the model in which to pick a goal, a procedure that we call Model Babbling. Once a model is chosen, it finds a goal in its sensory space, and infer motor parameters (that can be in the sensory space of a lower-level model) to reach that goal. Then, it passes those parameters as a goal to a lower-level model, which similarly infers motor parameters and passes those ones until the actual *Arm* motor space gets parameters to execute in the environment (with the same exploration noise as in Flat architectures). Model 4 has also to choose which lower-level model to use in order to reach an object end position s_o in S_{Object}, as two models (3 and 6) have S_{Object} as sensory space. Model 4 chooses the tool that enabled reaching s_o as close as possible in the past, e.g. if model 3 has in its history a reached sensory point s closer to s_o than any reached point with model 6, then model 3 is chosen. Finally, when motor parameters m are executed in the environment and feedback s is received, the mappings of all models are updated. However, only the tool-particular models are updated when a tool is currently held.

Random vs Active Model Babbling In a first condition, the agent randomly chooses the model that will find a goal, which is called Random Model Babbling (H-RMB). The problem of Model Babbling is an instance of strategic learning (Nguyen & Oudeyer, 2012), where different outcomes and strategies to learn them are available and the agent learns which strategies are useful for which outcomes. In that paper, they show that an active choice of the outcomes and strategies based on the learning progress on each of them increase learning efficiency compared to random choice. To develop active learning strategies, we first define a measure of learning progress for each of the 6 models. When a model has been chosen to babble, it draws a random goal s_g, and finds motor parameters m to reach it using the lower-level models. The actual outcome s in the sensory space of the model, associated to m might be very different from s_g as this goal might be unreachable, or because lower-level models are not mature enough for that goal. We define the competence associated to a goal s_g as the negative distance between the goal and the reached point, divided by the maximal distance in this space, to scale this measure across different spaces:

$$C(s_g) = -\frac{||s_g - s||}{max_{s'}||s' - s||} \quad (1)$$

and the interest $I(s_g)$ associated to this goal as

$$I(s_g) = |C(s_g) - mean_{kNN}C(s)| \quad (2)$$

where $mean_{kNN}C(s)$ is the mean competence of the ($k = 20$) nearest previous goals (k-Nearest Neighbours algorithm).

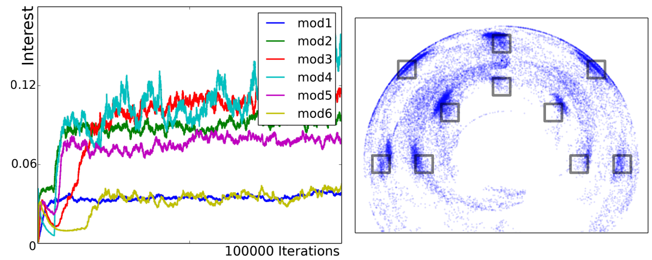


Figure 3: Condition H-P-AMB. Left: Interest of each model. Right: Exploration of the object space: each dot is the position reached with the object at the end of a movement.

The interest of a model is initialized at 0 and updated to follow the interest of the goals (with rate $n = 200$):

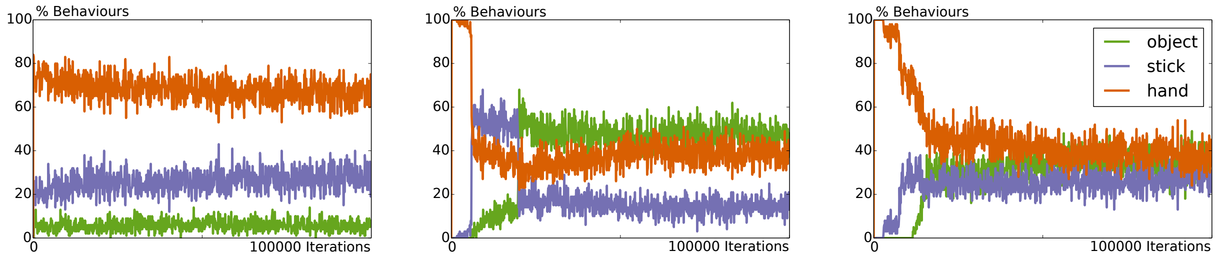
$$I_{model} = \frac{n-1}{n} I_{model} + \frac{1}{n} I(s_g) \quad (3)$$

In condition H-P-AMB, the choice of model is probabilistic and has $\varepsilon = 10\%$ chance to be random, and $(1 - \varepsilon)$ to be proportional to their interest. In condition H-GR-AMB, the choice of model is greedy (model with maximum interest) but also with $\varepsilon = 10\%$ of random choice. Finally, condition H-P-AMB-CTC (Curiosity-driven Tool Choice) is the same as H-P-AMB but the choice of the tool to use (model 3 or 6) is made with probabilities proportional to the interest of the two models, instead of being based on the more competent tool for the given object goal position. We call HACOB this Hierarchical Active Curiosity-driven mOdel Babbling algorithmic architecture with the algorithms H-P-AMB and H-P-AMB-CTC being two variants of the architecture.

Results

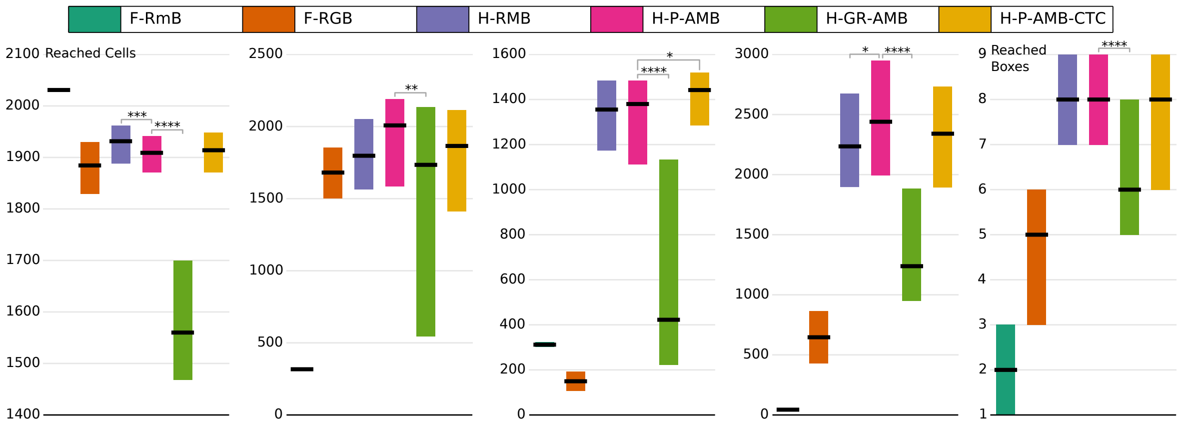
We perform 100 independent simulations of 100000 iterations per condition, starting with 100 iterations of motor babbling. Fig. 3 shows details about one trial of the condition H-P-AMB. We can see the interest of each model during one simulation, and the corresponding explored object space. The interests of models 2 and 5 increase once the arm succeeded to grab the corresponding stick. Following that, the interests of models 3 and 6 increase once the object has been reached.

Structure of the Evolution of Behaviours We provide a measure of three types of behaviours with objects during exploration. In the first category (*hand*) the arm did not grab any stick and thus did not move the out-of-reach object. In the second category (*stick*), the arm did grab one of the two sticks but did not touch the object with it. The third category (*object*) contains the movements where both a stick was grabbed and the object was moved by the stick. Fig. 4 shows a typical evolution of the proportion of the three categories of behaviours. We performed a more detailed analysis (see Table 1) by counting the trials where the evolution of the behaviours were similar to the ones found in infant development of the interaction with object (Guerin et al., 2013). A structure was considered similar to infant behavioral structures if



(a) No developmental structure (b) Developmental stages, abrupt changes (c) Overlapping phases structure

Figure 4: Typical behavioral evolution in the conditions (a) F-RGB, (b) H-GR-AMB, (c) H-P-AMB.



(a) *Hand* (b) $Stick_1$ (c) $Stick_2$ (d) *Object* (e) *Boxes*

Figure 5: Exploration of sensory spaces. Box plots show medians and quartiles of the 100 trials.

it validated each of the following criteria: behaviours of categories *stick* and *object* increase from 0 to more than 10% (potentially after an initial phase with a steady low value), are followed by a curve with small slope and no abrupt changes, and behaviours of category *object* start to raise at least 1000 iterations after *stick* started to raise (see Fig. 4(c) for a valid instance). Also, the median number of abrupt changes across trials are reported in Table 1 (as the sum of steady changes of more than 10% in the three behaviours), with a significant difference between condition H-GR-AMB and others (Mann-Whitney U tests, $p < 10^{-4}$).

Table 1: Behavioural analysis

Condition	Number of Trials validating criteria	Median number of Abrupt changes
F-RmB	0	0
F-RGB	0	1
H-RMB	60	2
H-P-AMB	70	2
H-GR-AMB	7	6
H-P-AMB-CTC	79	1

Exploration Efficiency Also, for each condition we measured the total exploration of the sensory spaces during training. The exploration of the hand, sticks and object spaces is defined as the number of reached cells in a 100×100 discretization of the (X,Y) space of their position at the end of the movement. Boxes' exploration is the number of boxes reached with the object during training. Fig. 5 shows the exploration of the different sensory spaces for each condition. We provide Mann-Whitney U test results of comparisons of total exploration for some pairs of conditions. One star means $p < 0.05$, two: $p < 10^{-2}$, three: $p < 10^{-3}$, four: $p < 10^{-4}$.

Structure of Tool Choice Finally, we compare the structure of tool choice made to reach object goal positions in two conditions for which only this choice differs. Fig. 6 shows the choice of tool to reach a given object goal position in the conditions H-P-AMB and H-P-AMB-CTC. When model 4 is babbling, it infers the best object position s_o to reach a random goal $s_b \in S_{Boxes}$. We plot all the choices that model 4 made during exploration, at position s_o on a 2D space, with color blue if $Stick_1$ was chosen and red if $Stick_2$ was chosen. In condition H-P-AMB, we can see strong boundaries between tool choice regions. By contrast, in condition H-P-AMB-CTC, both tools are chosen in all regions.

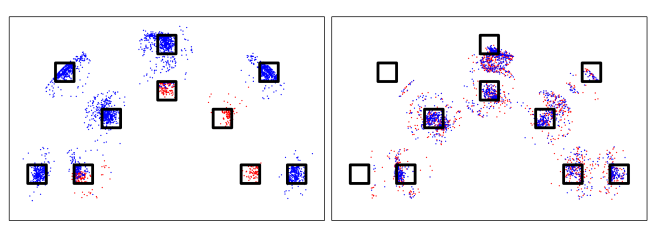


Figure 6: Chosen tool depending on object goal position. Blue: long stick choice. Red: small stick. Left: H-P-AMB, strong boundaries between tool choice regions. Right: H-P-AMB-CTC, parallel exploration of both tools in all regions.

Discussion

Structure of the Evolution of Behaviours Results show different structures of behaviour evolution in the different conditions. Flat architectures cannot efficiently learn in this environment with a high-dimensional sensory space. Therefore, they do not show structure in the behavioural evolution but rather steady proportions of the three behaviours. By contrast, hierarchical condition H-GR-AMB shows successive behavioural steps with abrupt changes, which reflects the greedy choice of model to babble. When one model becomes more interesting than another, it is chosen for a large number of iterations until another model exceeds its interest. Random model babbling shows overlapping phases structures more compatible with infants' studies in the evolution of the three behaviours, but less than active model babbling (60% instead of 70% or 79%). This is because random model babbling does not adapt its choice of models to their interests along development. Indeed, it often explores model 1 even if it is sufficiently explored to make progress on higher-level models, and so explores less the object position space than active model babbling (H-P-AMB). Also, all models are still useful to explore after the number of iterations simulated here so the first behavioural phases (*hand* and *stick*) do not lessen towards the end of the simulations in condition H-P-AMB.

A notable difference between active and random model babbling is on the cognitive or intentional level, as active model babbling monitors the current progress of each model whereas random model babbling does not. Furthermore, in other setups where some tasks are learned much faster than others and where at some point it becomes useless to explore a mastered task, active model babbling should also show an important difference both on a quantitative exploration point of view and on the structure of the evolution of behaviours.

Variability of Strategies to Reach Goals The comparison of conditions H-P-AMB and H-P-AMB-CTC shows that when the agent chooses the tool to reach a given object goal position based on the interest of the corresponding models, both tools are trained to reach all goals instead of training only the best performing tool. Indeed, with the active curiosity-driven choice of tool, the small stick has produced more diverse effects on the object than in the optimal tool's condition (Fig. 5c), even if those effects could also have been generated with the long tool. Curiosity-driven active tool choice shows more child-like results in accordance with Siegler's overlapping waves theory, which describes the use of strategies in infants and explains that non-optimal strategies are also explored as they might turn out to be finally good ones for this problem or for different but related ones.

Finally, we first demonstrated that hierarchical structure is a determining factor for the emergence of structured behavioural phases in our simple tool use setup. Then we showed that the active exploration of this hierarchical structure with curiosity-driven mechanisms combined with goal babbling reinforces this emergence and is essential to efficiently learn in this hierarchy. Although mechanisms such as action observation, sequential learning or causal inference are known to be highly important mechanisms in human development, we thus suggest that curiosity-driven exploration and goal babbling should also be considered as ones of them, but they have comparatively little been studied so far.

References

Baldassarre, G., & Mirolli, M. (2013). *Intrinsically motivated learning in natural and artificial systems*. Springer.

Baranes, A., & Oudeyer, P.-Y. (2013). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, *61*(1).

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11).

Guerin, F., Kruger, N., & Kraft, D. (2013). A survey of the ontogeny of tool use: from sensorimotor experience to planning. *Autonomous Mental Development, IEEE Transactions on*, *5*(1).

Ijspeert, A. J., Nakanishi, J., Hoffmann, H., Pastor, P., & Schaal, S. (2013). Dynamical movement primitives: learning attractor models for motor behaviors. *Neural computation*, *25*(2).

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460.

Moulin-Frier, C., Nguyen, S. M., & Oudeyer, P.-Y. (2014). Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Frontiers in Psychology*, *4*.

Nguyen, S., & Oudeyer, P.-Y. (2012). Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn*, *3*(3).

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, *11*(2).

Piaget, J., Cook, M., & Norton, W. (1952). *The origins of intelligence in children* (Vol. 8) (No. 5). International Universities Press New York.

Rolf, M., Steil, J., & Gienger, M. (2010). Goal babbling permits direct learning of inverse kinematics. *Autonomous Mental Development, IEEE Transactions on*, *2*(3).

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford University Press.

Stoytchev, A. (2005). Behavior-grounded representation of tool affordances. In *Proceedings of the 2005 ieee international conference on robotics and automation.*

Ugur, E., Nagai, Y., Sahin, E., & Oztop, E. (2015). Staged development of robot skills: Behavior formation, affordance learning and imitation with motionese. *IEEE Transactions on Autonomous Mental Development*, *7*(2).

Zelazo, P., & Kearsley, R. (1980). The emergence of functional play in infants: Evidence for a major cognitive transition. *Journal of Applied Developmental Psychology*, *1*(2).