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

Sébastien Forestier (sebastien.forestier@inria.fr)
INRIA Bordeaux Sud-Ouest
Bordeaux, France

Pierre-Yves Oudeyer (pierre-yves.oudeyer@inria.fr)
INRIA Bordeaux Sud-Ouest
Bordeaux, France

Abstract

This is the abstract.

Keywords: curiosity-driven learning; tool use; goal babbling; overlapping waves;

Introduction

Development of tool use: different properties (Guerin, Kruger, & Kraft, 2013) (cite others?). List of some properties: grounding of representation and planning based on a large amount of experiences, ongoing process of upgrading representations (others?). In that paper we will focus on one important property in the development of precursors of tool use which is the seamless progression between successives phases of behavior with tools and objects which are called overlapping waves [Siegler]. One behavioral description of the successives phases differentiate three phases (Guerin et al., 2013). In the first phase the babies engage mostly in behavior without objects (babbling), in the second phase their behavior has shifted towards interaction with one object, and then interaction between objects. [more details: months, experiments]

We hypothesize that several mechanisms play a role in the structure of this behavioral progression and in particular 1) the intrinsic motivation to explore as a self-regulation of the learning growth of complexity, and 2) the structure of the representation used to encode sensorimotor experience.

Curiosity studies in developmental psychology (Kidd, Piantadosi, & Aslin, 2012) (Gottlieb, Oudeyer, Lopes, & Baranes, 2013)

We will study aspects of these hypothesis leveraging and extending models of curiosity-driven learning of sensorimotor models to the exploration of given hierarchies of sensorimotor models. In such hierarchies, parts of the sensory space (e.g. the position of the hand) can be used as a motor space by another higher-level model to explore other sensory spaces. We do not study some other important factors of the development of tool use: the autonomous building and evolution of the hierarchy of models but we consider it given to a learning agent. We also do not address the question of the role of social guidance.

Related work Curiosity-driven modelling work, emergence of developmental trajectories. (Oudeyer, Kaplan, & Hafner, 2007) (Oudeyer, 2007) (Csikszentmihalyi, 1990) (Schmidhuber, 1991) (Santucci, Baldassarre, & Mirolli, 2013) (Cangelosi et al., 2010) (Oudeyer & Smith, 2014)

IAC series of architectures and Explauto framework: previous experiments. (Moulin-Frier, Nguyen, & Oudeyer, 2014) (Moulin-Frier, Rouanet, Oudeyer, & others, 2014) (Baranes & Oudeyer, 2010) (Baranes & Oudeyer, 2009) (Baranes & Oudeyer, 2013)

Representations in explauto and other models (Mugan & Kuipers, 2009a) (Metzen & Kirchner, 2013) (Sutton et al., 2011) (Mugan & Kuipers, 2009b) (Vigorito & Barto, 2010) (Sutton, Precup, & Singh, 1999)

Other related work (Ugur, Nagai, Sahin, & Oztop, 2015) (Schmerling, Schillaci, & Hafner, 2015) (Forestier & Oudeyer, 2015) (Sánchez-Fibla, Forestier, Ysard, Moulin-Frier, & Verschure, 2016)

More details on experiments (Ijspeert, Nakanishi, Hoffmann, Pastor, & Schaal, 2013)

()

Along with this paper we provide open-source Python code[1] with iPython/Jupyter notebooks that explain how to reproduce the experiments and analysis.

Methods

Environment

We simulate a 2D robotic arm using tools to move an object into different boxes in the environment. In each trial, we execute a motor trajectory given by the agent, we evaluate its consequences on the sensory dimensions and we give him this sensory feedback. Finally the arm, tools and objects are resetted to their intial state.

The next sections precisely describe the different items of the environment and their interactions. See Fig.1 for an exemple of the state of the environment.

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 initial position, mapped to a standard interval of $[-1,1]$. The length of the 3 parts of the arm are 0.5, 0.3 and 0.2 so the total length of the arm is 1 unit. The initial position of the arm is vertical with each joint 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

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

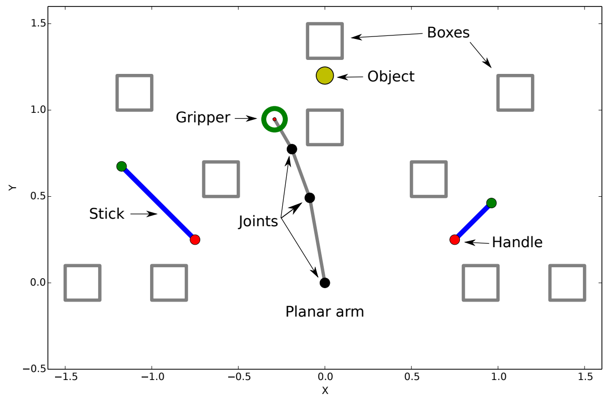


Figure 1: Play Environment

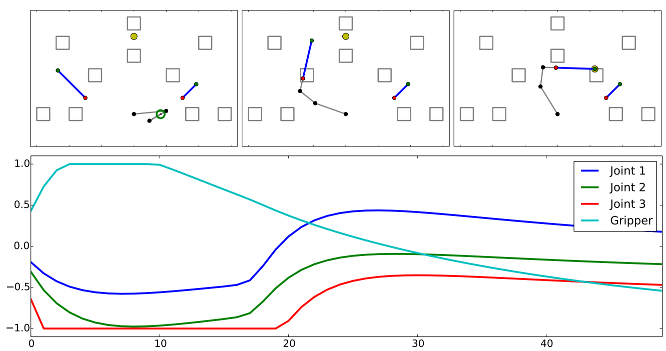


Figure 2: Exemple of one mouvement. Top: position of the arm at time steps 17, 33 and 50 along the 50 steps movement. Bottom: value of the four Dynamical Movement Primitives during the movement.

its initial position is *open* (with $g = 0$). The robotic arm thus 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 Primitive (Ijspeert et al., 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 with a starting and a goal position equal to the rest position of the joint. Each DMP is parameterized by one weight on each of 3 basis functions whose centers are distributed homogeneously throughout the movement duration. The weights are bounded in the interval $[-200,200]$ (mapped to the standard interval $[-1,1]$) which 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 by 12 weights which are represented by the motor space $M = [-1,1]^{12}$.

Objects and tools A yellow sphere can be moved into one of the 4 fixed squared boxes. The initial position of the sphere is $(0,1.2)$ and is thus unreachable directly with the gripper. One of two sticks can be grasped in order to reach the object. A small stick of length 0.3 is located on the right of the arm, with initial position $(0.75,0.25)$ and initial angle $\frac{\pi}{4}$ from the horizontal line. A long stick of length 0.6 is located on the left of the arm, with initial position $(-0.75,0.25)$ and initial angle $\frac{3\pi}{4}$ from the horizontal line as in Fig. 1. If the gripper closes near the end of one of the sticks (closer than 0.1), it is considered grasped and will follow the gripper's position and the angle of the arm's last part until the gripper opens. Similarly, if the other end of a stick reaches the sphere (within 0.1), the object will follow the end of the stick. Ten boxes have identifiers 1 to 10 and are static at positions $(-1.4,0)$, $(-1.1,1.1)$, $(0,1.4)$, $(1.1,1.1)$, $(1.4,0)$, $(-0.9,0)$, $(-0.6,0.6)$, $(0,0.9)$, $(0.6,0.6)$ and $(0.9,0)$ and have size 0.2. The first five boxes can only be reached with the long stick, and the other five can be reached by the two sticks. At the end of the trial, the object is considered to be in one of the box if its center is in the box.

Sensory feedback At the end of the movement, the robot gets sensory feedback from the different items of the environment. It gets the trajectory of its hand and gripper, the trajectory of the end of the sticks, the end position of the object, and whether the object is in each box. The trajectory of the hand and of the end point of the sticks are represented by sequences of x and y positions at different time points: steps 12, 25, 37 during the movement of 50 steps (6D for the hand and for each stick). Similarly, the trajectory of the gripper is a sequence of 1 or -1 depending whether the gripper is open or closed (3D). The agent receives the identifier of the reached box if one of them has been reached, 0 otherwise. He also gets the minimal distance of the object (at the end of the movement) to the center a box, even if none have been reached. The sensory information thus contains 9 values for the trajectory of the hand and gripper (S_{Hand}), 6 for the trajectory of the end of each stick (S_{Stick_1} and S_{Stick_2}), 2 for the end position of the object (S_{Object}) and 2 for the boxes (S_{Boxes}). The sensory space has a total of 25 dimensions.

Learning architectures

We describe in this section the different learning architectures used in the experiment.

Explauto framework We use the Explauto autonomous exploration library (Moulin-Frier, Rouanet, et al., 2014) to easily define experiments with robots exploring their environment. In this framework, the agent explores a mapping between its given motor space M and sensory space S, using a sensorimotor model that learns the mapping and an interest model that chooses which regions of the spaces to explore. The sensorimotor model processes new information of the

Figure 3: Flat architecture

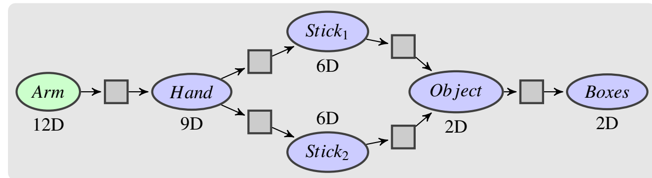


Figure 4: Hierarchy of sensorimotor models

form (m, s) with $m \in M$ being the experimented motor parameters and $s \in S$ the received sensory feedback corresponding to that parameters. It provides forward predictions of probable s given m and inverse inference of a probable m to reach a given s. We use the simplest sensorimotor model, the nearest neighbor algorithm that performs the prediction of s given m using the nearest neighbor of m in M in the database of the previous experiments and returning its sensory part. The inverse inference is computed as the motor part of the nearest neighbor in S of the given s, but with some exploration noise to allow novel motor parameters to be explored.

The agent also needs an interest model that chooses which regions of the motor or sensory space to explore. If the agent chooses goals in the motor space and explores them it is called Motor Babbling, and if it chooses the goals to explore in the sensory space, this is Goal Babbling. To choose the goals in this space of interest, different strategies are possible. The simplest one is to draw random goals in this space, but strategies based on the learning progress in different regions of the space have been shown more efficient in [ref]. This strategies are called active as the agent autonomously drives its exploration towards regions of the space that are both reachable and learnable. We use the SAGG-RIAC architecture (Baranes & Oudeyer, 2013) where the space of interest is the sensory space (Goal Babbling) and that incrementally splits this space into subregions where the learning progress is different. The learning progress is here computed as the absolute derivative of the reaching competence. This competence associated to a sensory goal is minus the distance between the goal and the sensory point.

Flat architecture

Hierarchical architecture We present here an architecture that represents sensorimotor information with a hierarchical structure in Fig. 4.

Only the motor module (mod1) adds exploration noise ($\sigma = 0.02$) even in hierarchical architecture. That was the key to have more efficient hierarchical exploration. Indeed, if all modules successively add exploration noise, few iterations succeed in touching the object. Alternatively, if the exploration noise is reduced, exploration is less efficient as in NN only the motor module will finally apply noise on known motor commands. With regression instead of NN, noise can instead be put only on the babbling module.

How competence and interest of modules is computed.

Experiments

NN, 100 iterations of Motor babbling and then 300000 iterations of the condition. 10 trials per condition, mean and std provided.

Conditions

- F-RmB: Random Motor Babbling with a flat architecture learning $M \rightarrow S_{Hand} \times S_{Stick_1} \times S_{Stick_2} \times S_{Object} \times S_{Boxes}$
- F-RGB: Flat architecture, Random Goal Babbling
- F-AGB: The same architecture but with active goal babbling (my version of SAGG-RIAC)
- H-RGB-P-AMB: Hierarchical architecture with Random Goal Babbling in each module, and choice of module that babbles based on interest (ε-prop: probabilities proportional to interest but with $\varepsilon = 10\%$ of random choice). Choice of tool to use based on the maximum competence of the two modules to reach the goal object point.
- H-RGB-GR-AMB: same as H-RGB-P-AMB but the choice of module to babble is ε-greedy with $\varepsilon = 0.1$
- H-RGB-P-AMB-PGITC: same as H-RGB-P-AMB but the choice of the tool to use is based on the interest (global) of the two modules (ε-prop).

Measures Exploration of the different sensory spaces (number of reached cells in a discretization of the space divided by number of cells).

Results

We measure the exploration of the different sensory spaces in the 6 conditions each 5000 iterations along the 300000 iterations (Fig. 5). We provide statistical Mann-Whitney U test results of comparisons of the exploration at the end of the experiments in different pairs of conditions. We first show that the Motor Babbling condition (H0-MB) quickly explores S_{Hand} but explores badly the other sensory spaces, compared to the other conditions (STATS). Then H0-GB and H0-Tr have similar exploration results (STATS maybe Tr better in Boxes). Then H1-GR shows lower exploration of S_{Hand}, S_{Stick_1} and S_{Stick_2} than H0-GB and H0-Tr but higher exploration of S_{Object} and S_{Boxes} (STATS). Also, H1-GR shows lower exploration of all spaces than H1 (STATS). H1 shows lower exploration of S_{Hand} than H0-GB and H0-Tr (STATS), similar (STATS?) exploration of S_{Stick_1} and S_{Stick_2} but higher exploration of S_{Object} and S_{Boxes} (STATS). H1-CL vs H1 (STATS?).

Fig. 6 shows more details about one (randomly chosen) trial of the condition H1. We can see the interest of each module during the whole experiment. The interests of modules 1, 2, 5 are almost stable, as the spaces of interests are high-dimensional and homogeneously reachable: a random exploration will find regions to learn more for more iterations than 300000. The interests of modules 3 and 6 increase abruptly when the object is touch by the corresponding tool. An exemple of exploration of the 2D space with the object is also provided in Fig. 6(b) also corresponding to condition H1.

Fig. 8 shows a comparison of the choice of tool to reach a given object goal position in the conditions H1 and H1-CL. In those conditions, module 4 learns a mapping between S_{Object} and S_{Boxes}. When this module is babbling (give # babbling in the 2 conditions: around 100000), it chooses a random goal $s_b \in S_{Boxes}$ and infers the best object position s_o to reach s_b. To reach s_o, one of the tools ($Stick_1$ with module 3 or $Stick_2$ with module 6) has to be chosen. We plot all those choices, at position s_o on a 2D space, with color blue if $Stick_1$ was chosen and red if $Stick_2$ was chosen, with one plot for each of the two conditions. We see (or soon will see in new plots, and maybe with smoothed stats along trials?) that in both conditions, goals that can only be reached with the long stick (the furthest) are actually chosen to be explored with the long stick. However, goal that can be reached with both tools are more often chosen to be explored with the long stick in the interest-based choice of condition H1 than in competence-based choice of condition H1-CL.

Discussion

F-AGB vs H-RGB-P-AMB vs H-RGB-GR-AMB

H-RGB-P-AMB vs H-RGB-P-AMB-PGITC

Conclusion

Acknowledgments

References

Baranes, A., & Oudeyer, P.-Y. (2009). R-iac: Robust intrinsically motivated exploration and active learning. *Autonomous Mental Development, IEEE Transactions on*, *1*(3), 155–169.

Baranes, A., & Oudeyer, P.-Y. (2010). Intrinsically motivated goal exploration for active motor learning in robots: A case study. In *Intelligent robots and systems (iros), 2010 ieee/rsj international conference on* (pp. 1766–1773).

Baranes, A., & Oudeyer, P.-Y. (2013, January). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, *61*(1), 49–73. doi: 10.1016/j.robot.2012.05.008

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., ... others (2010). Integration of action and language knowledge: A roadmap for developmental robotics. *Autonomous Mental Development, IEEE Transactions on*, *2*(3), 167–195.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row. Retrieved from `http://books.google.fr/books?id=V9KrQgAACAAJ`

Forestier, S., & Oudeyer, P.-Y. (2015). Towards hierarchical curiosity-driven exploration of sensorimotor models. In *2015 joint ieee international conference on development and learning and epigenetic robotics (icdl-epirob)* (p. 234-235).

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013, November). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585–593. doi: 10.1016/j.tics.2013.09.001

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), 18–45.

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), 328–373.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, *7*(5), e36399.

Metzen, J. H., & Kirchner, F. (2013). Incremental learning of skill collections based on intrinsic motivation. *Frontiers in Neurorobotics*, *7*. doi: 10.3389/fnbot.2013.00011

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*.

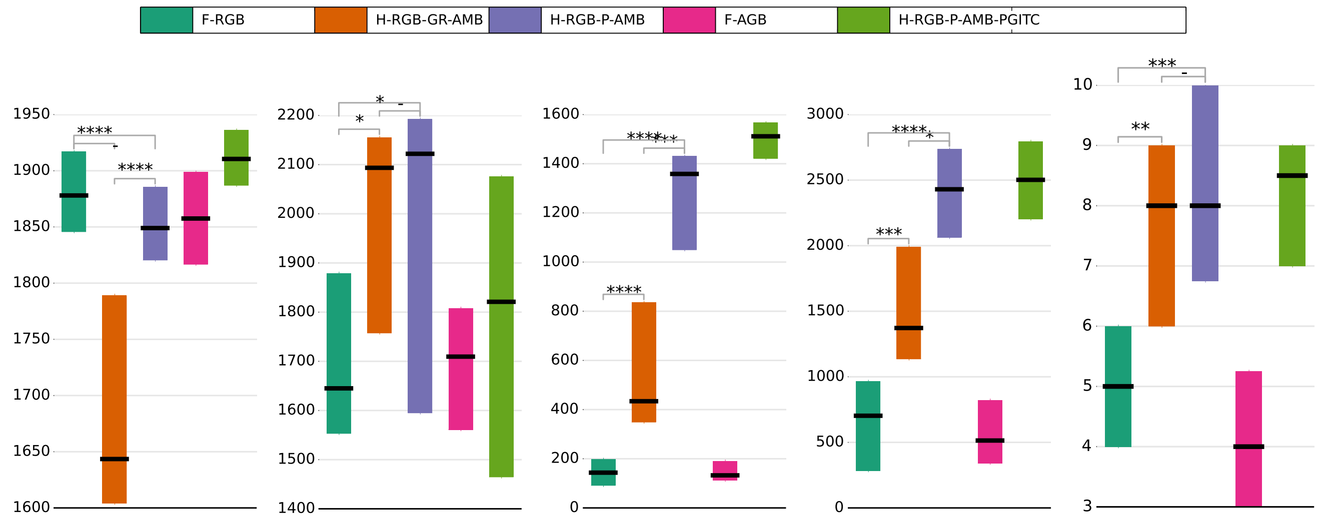
Moulin-Frier, C., Rouanet, P., Oudeyer, P.-Y., & others. (2014). Explauto: an open-source Python library to study autonomous exploration in developmental robotics. In *ICDL-Epirob-International Conference on Development and Learning, Epirob.*

Mugan, J., & Kuipers, B. (2009a). Autonomously Learning an Action Hierarchy Using a Learned Qualitative State Representation. In *IJCAI* (pp. 1175–1180).

Mugan, J., & Kuipers, B. (2009b). Autonomously learning an action hierarchy using a learned qualitative state representation.

Oudeyer, P.-Y. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, *1*. doi: 10.3389/neuro.12.006.2007

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007, April). Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, *11*(2), 265–286.



(b) *Hand* (c) *Stick*$_1$ (d) *Stick*$_2$ (e) *Object* (f) *Boxes*

Figure 5: Exploration of sensory spaces.

Oudeyer, P.-Y., & Smith, L. (2014). How evolution may work through curiosity-driven developmental process.

Sánchez-Fibla, M., Forestier, S., Ysard, J., Moulin-Frier, C., & Verschure, P. (2016). Unifying affordance and kinematics learning: a computational approach to bimanual affordances. *submitted*.

Santucci, V. G., Baldassarre, G., & Mirolli, M. (2013). Which is the best intrinsic motivation signal for learning multiple skills? *Frontiers in Neurorobotics*, *7*. doi: 10.3389/fnbot.2013.00022

Schmerling, M., Schillaci, G., & Hafner, V. V. (2015). Goal-directed learning of hand-eye coordination in a humanoid robot. In *5th joint ieee international conferences on development and learning and on epigenetic robotics (icdl-epirob).*

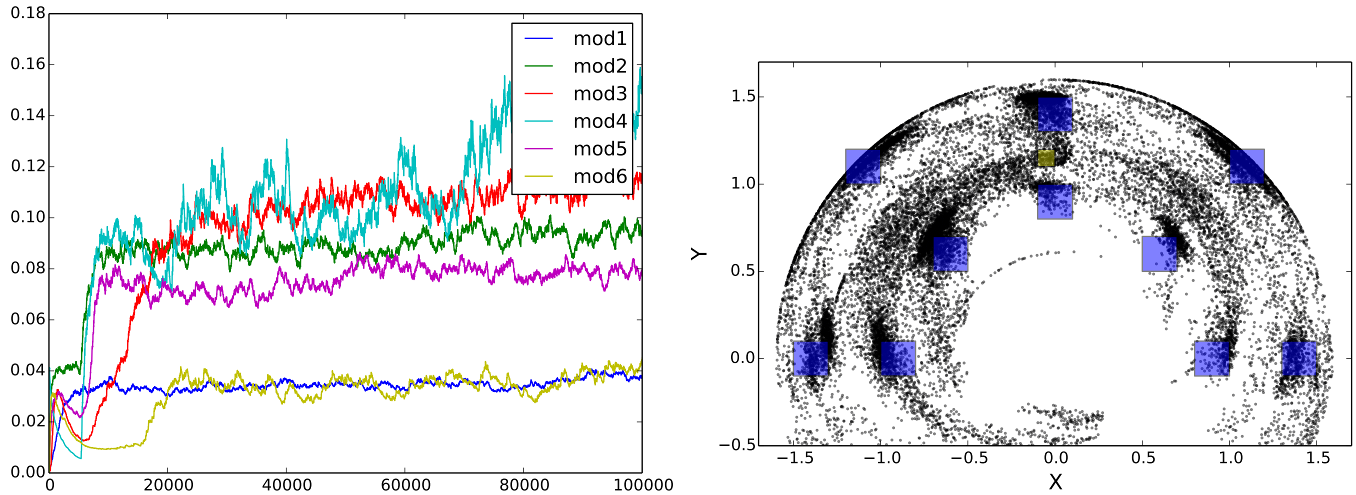
Schmidhuber, J. (1991). A possibility for Implementing curiosity and boredom in model-building neural controllers.

Sutton, R. S., Modayil, J., Delp, M., Degris, T., Pilarski, P. M., White, A., & Precup, D. (2011). Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *The 10th international conference on autonomous agents and multiagent systems-volume 2* (pp. 761–768).

Sutton, R. S., Precup, D., & Singh, S. (1999). Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, *112*(1), 181–211.

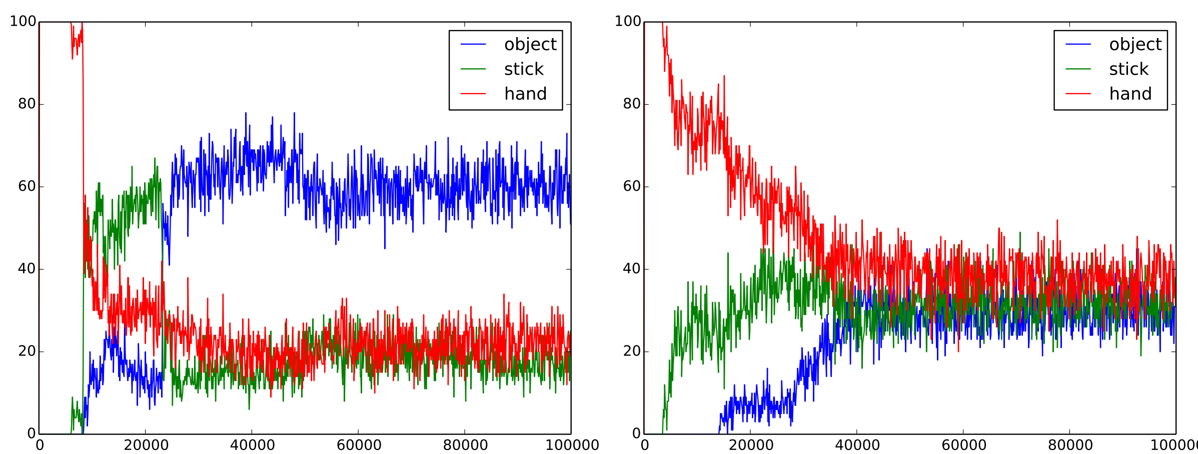
Ugur, E., Nagai, Y., Sahin, E., & Oztop, E. (2015, June). Staged development of robot skills: Behavior formation, affordance learning and imitation with motionese. *Autonomous Mental Development, IEEE Transactions on*, *7*(2), 119-139. doi: 10.1109/TAMD.2015.2426192

Vigorito, C. M., & Barto, A. G. (2010). Intrinsically motivated hierarchical skill learning in structured environments. *Autonomous Mental Development, IEEE Transactions on*, *2*(2), 132–143.



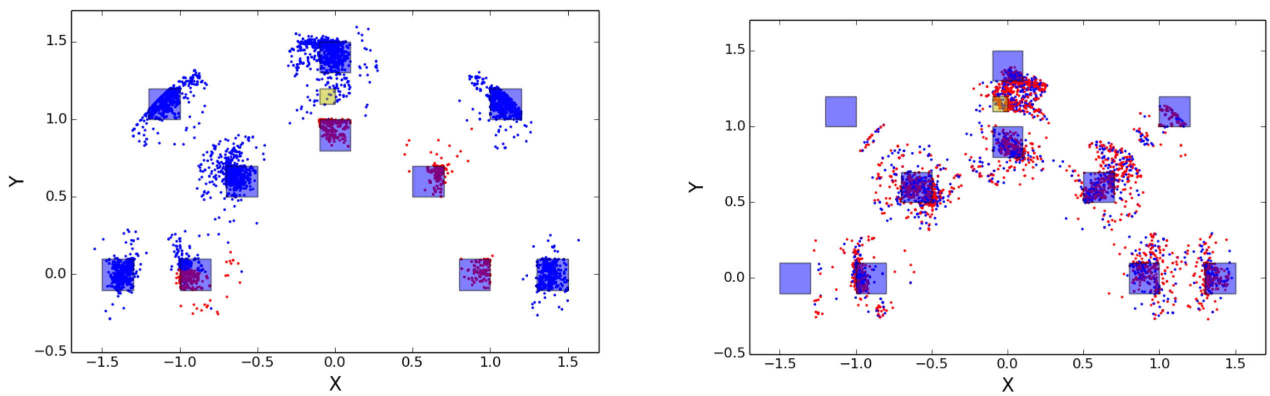
(a) (b)

Figure 6: Condition H1. (a) Interests of each module. (b) Exploration of the object space: each dot is one point reached with the object at the end of one movement.



(a) H-RGB-GR-AMB (b) H-RGB-P-AMB

Figure 7: Behavioral measure



(a) H-RGB-P-AMB

(b) H-RGB-P-AMB-PGITC

Figure 8: Chosen tool depending on object goal position. Blue points: long stick choice. Red points: small stick choice.