

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 successive phases of behavior with tools and objects which are called overlapping waves [Siegler]. One behavioral description of the successive 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, 2007) (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, 2013)

Representations in explauto and other models (Mugan & Kuipers, 2009) (Metzen & Kirchner, 2013) (Vigorito & Barto, 2010) (Sutton, Precup, & Singh, 1999)

Other related work (Ugur, Nagai, Sahin, & Oztog, 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 code¹ 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 reset to their initial state. The next sections precisely describe the different items of the environment and their interactions. See Fig.1 for an exemple 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 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 total length of the arm is 1 unit. The resting 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 its resting 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

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

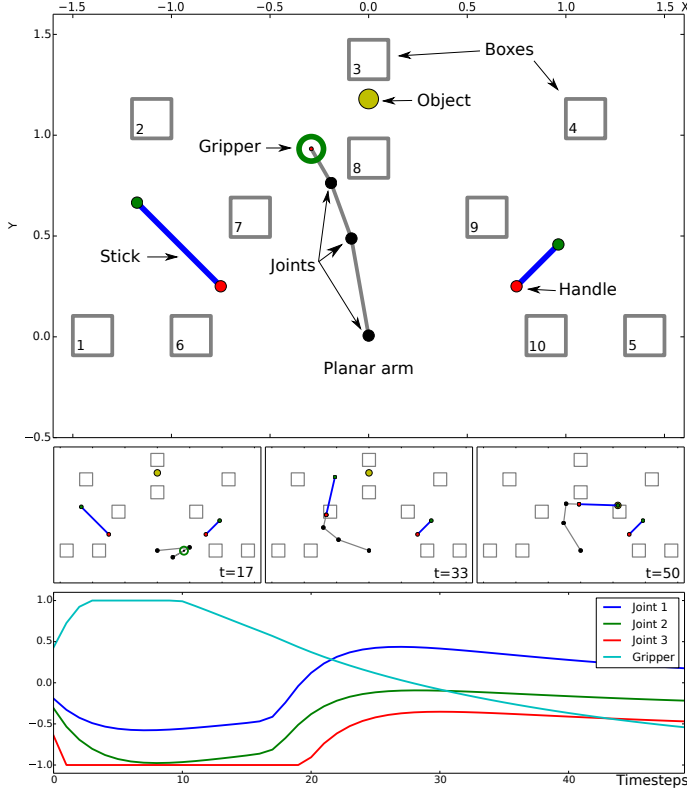


Figure 1: Top: a state of the environment. Middle: position of the arm at time steps 17, 33 and 50 along the 50 steps movement. Bottom: value of the four DMPs during the movement.

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 ten fixed squared boxes. The initial position of the sphere is $(0, 1.2)$ and is thus unreachable directly with the gripper. Two sticks can be grasped in order to reach the object. A small stick of length 0.3 is located at position $(0.75, 0.25)$ with initial angle $\frac{\pi}{4}$ from the horizontal line. A long stick of length 0.6 is located at position $(-0.75, 0.25)$ with initial angle $\frac{3\pi}{4}$ as in Fig. 1. If the gripper is closed near the handle of one stick (closer than 0.2), it is considered grasped and

will follow the gripper’s position and the angle of the arm’s last segment until the gripper opens. Similarly, if the other end of a stick reaches the sphere (within 0.1), the sphere will follow the end of the stick. 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. At the end of the movement, 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 the gripper (S_{Hand}), the trajectory of the end of the sticks (S_{Stick_1} and S_{Stick_2}), the position of the object at the end of the movement (S_{Object}), and whether the object is in a box at the end of the movement and the distance between the object and the nearest box (S_{Boxes}). 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 (9D). Similarly, the trajectories of the end points of the sticks are 3-points sequences of x and y positions (6D for each stick). The agent receives the identifier (from 1 to 10) of the reached box if one of them has been reached by the sphere, 0 otherwise. He also gets the minimal distance of the object (at the end of the movement) to the center of a box, even if none have been reached. The sensory information thus contains 9 values for the trajectory of the gripper, 6 for the trajectory of the end of each stick, 2 for the end position of the object and 2 for the boxes. The total sensory space S has 25 dimensions.

Learning architectures

The problem settings for the learning agent is to explore its sensorimotor space by iteratively choosing motor parameters that represents arm trajectories and receiving sensory feedback. In this section we describe the different learning architectures that we will compare in the experiment.

Flat architectures We define a flat architecture as directly learning a mapping between the motor space M (12D) and the sensory space S (25D). The control condition is a random motor babbling condition (F-RmB) that randomly chooses new motor parameters m to try at each iteration. In the other conditions, the agent performs Goal Babbling, by self-generating a goal in the sensory space at each iteration and trying to reach it. 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 neighbor algorithm: it gets the motor part of the nearest neighbor of the given s in S , and adds exploration noise (gaussian with $\sigma = 0.01$) to allow new motor parameters to be explored.

The agent also needs an interest model that chooses goals in the sensory space. To generate those goals, different

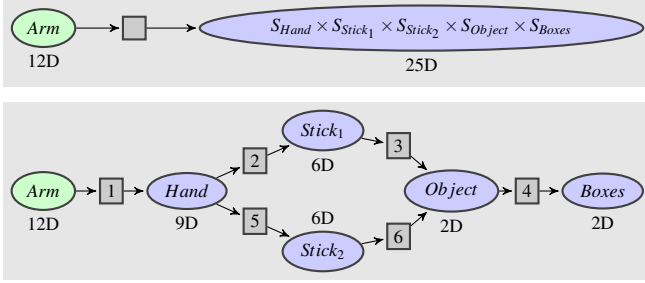


Figure 2: Architectures. Top: Flat. Bottom: Hierarchical.

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 idea can't be applied in a 25D sensory space as a learning progress signal can't be properly estimated in this volume. Thus we use a simpler random generation of goals in the sensory space as an interest model in the flat random goal babbling condition (F-RGB). We use the Explauto autonomous exploration library (Moulin-Frier, Rouanet, et al., 2014) to easily define those sensorimotor and interest models.

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 influence the gripper, which influence one of the tools (but not both at the same time), which influence 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 have to choose the model that will choose 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 a sensory space) to reach that goal. Then, it passes those parameters as a goal to be reached by the lower-level model, which similarly infers motor parameters and passes those ones until the actual *Arm* motor space gets parameters to try in the environment (with the same exploration noise as in Flat architectures). Model 4 have 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 have allowed to reached s_o as close as possible in the past. Finally, when motor parameters m have been tested in the environment and feedback s received, the mappings of all models are updated, but if model 4 chose one tool then the mapping of

the other tool is not.

A first condition is to randomly choose the model that will find a goal, this is Random Model Babbling (H-RGB-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 module, associated to m might be very different from s_g as this goal might be unreachable, or because lower-level modules are not mature enough for that goal. We define the competence associated to a goal s_g as $C(s_g) = -||s_g - s||$, and the interest $I(s_g)$ associated to this goal as the absolute difference between $C(s_g)$ and the mean competence of the ($k = 20$) nearest previous goals. The interest of a model is initialized at 0 and updated to follow the interest of the goals:

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

In condition H-RGB-P-AMB, the choice of module is proportional to their interest (probabilities are proportional to interest but with $\epsilon = 10\%$ of random choice). In condition H-RGB-GR-AMB, the choice of module is greedy (module with maximum interest) but also with $\epsilon = 10\%$ of random choice. Finally, condition H-RGB-P-AMB-PGITC is the same as H-RGB-P-AMB but the choice of the tool to use (model 3 or 6) is 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.

Results

Nearest Neighbors, 100 iterations of Motor babbling and then 100000 iterations of the condition. 100 trials per condition.

Measures

We provide a measure of the different types of behaviors with the sticks and the object during exploration. We categorize the behaviors into three types. In the first category (*hand*) are movements of the arm that did not grab any stick and thus not moved the out-of-reached object. The second category (*stick*) are movements that 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.

Also, for each condition we measure the total exploration of the different 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 the position at timestep 50 (end of movement).

The exploration of the boxes is the number of boxes that have been filled with the object during training.

Fig. 3 shows details about one trial of the condition H-RGB-P-AMB. We can see the interest of each module during the whole experiment. The interests of modules 2 and 5 increase abruptly once the arm succeeded in grabbing the corresponding stick. Following that, the interests of modules 3 and 6 also increase abruptly when the object has been touched by the corresponding stick. An example of exploration of the 2D space with the object is also provided in Fig. 3(b) corresponding to the same condition.

Fig. 4 shows the proportion of the three categories of behavior along the 100000 iterations for conditions H-RGB-GR-AMB and H-RGB-P-AMB.

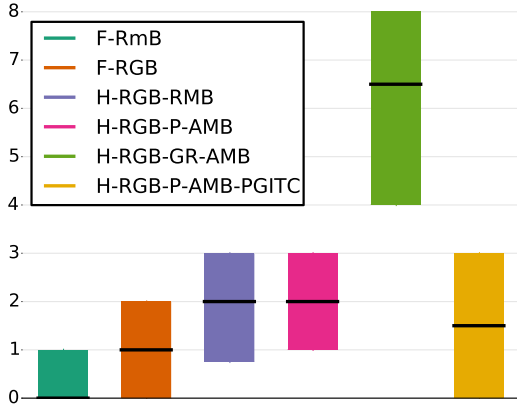


Figure 5: Number of abrupt behavioral changes

Fig. 6 shows the total exploration of the different sensory spaces for each condition. We provide statistical Mann-Whitney U test results of comparisons of the exploration at the end of the experiments in different pairs of conditions. Firstly, the Motor Babbling condition (F-RmB) have more explored S_{Hand} and less S_{Object} and S_{Boxes} compared to the other conditions ($p < 0.0001$). Then, F-RGB explores all spaces less than H-RGB-RMB condition ($p < 0.01$). Also, H-RGB-GR-AMB shows lower exploration all spaces than H-RGB-P-AMB ($p < 0.01$). Condition H-RGB-P-AMB-PGITC explores more S_{Stick_2} ($p < 0.05$) than condition H-RGB-P-AMB, and difference is not significant in other spaces.

Fig. 7 shows a comparison of the choice of tool to reach a given object goal position in the conditions H-RGB-P-AMB and H-RGB-P-AMB-PGITC. In those conditions, module 4 learns a mapping between S_{Object} and S_{Boxes} . When this module is babbling, 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) is 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, in one figure for each of the two conditions. We can see two very different choice structures. 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. (2013, January). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, 61(1).
- 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).
- 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)*.
- 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).
- 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., 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).
- Metzen, J. H., & Kirchner, F. (2013). Incremental learning of skill collections based on intrinsic motivation. *Frontiers in Neurobotics*, 7.
- 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.

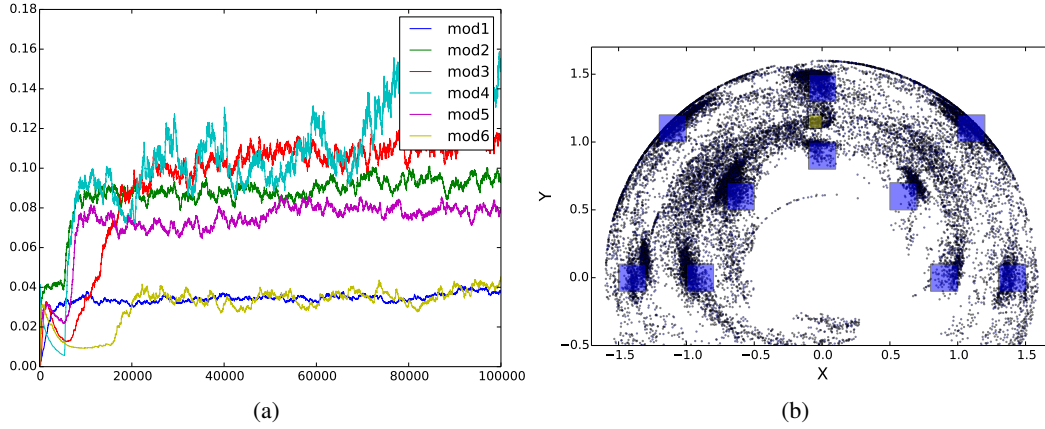


Figure 3: Condition H-RGB-P-AMB. (a) Interests of each module. (b) Exploration of the object space: each dot is one point reached with the object at the end of one movement.

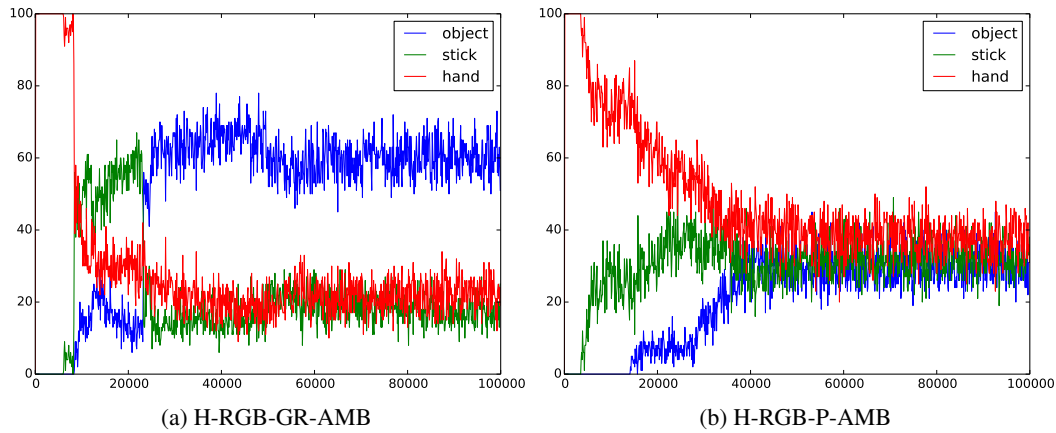


Figure 4: Behavioral measure

- In *ICDL-Epirob-International Conference on Development and Learning, Epirob*.
- Mugan, J., & Kuipers, B. (2009). Autonomously Learning an Action Hierarchy Using a Learned Qualitative State Representation. In *IJCAI*.
- 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. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neuro-robotics*, 1.
- 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 Neuro-robotics*, 7.
- 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-epi-rob)*.
- Schmidhuber, J. (1991). A possibility for Implementing curiosity and boredom in model-building neural controllers.
- 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).
- 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).
- Vigorito, C. M., & Barto, A. G. (2010). Intrinsically motivated hierarchical skill learning in structured environments. *Autonomous Mental Development, IEEE Transactions on*, 2(2).

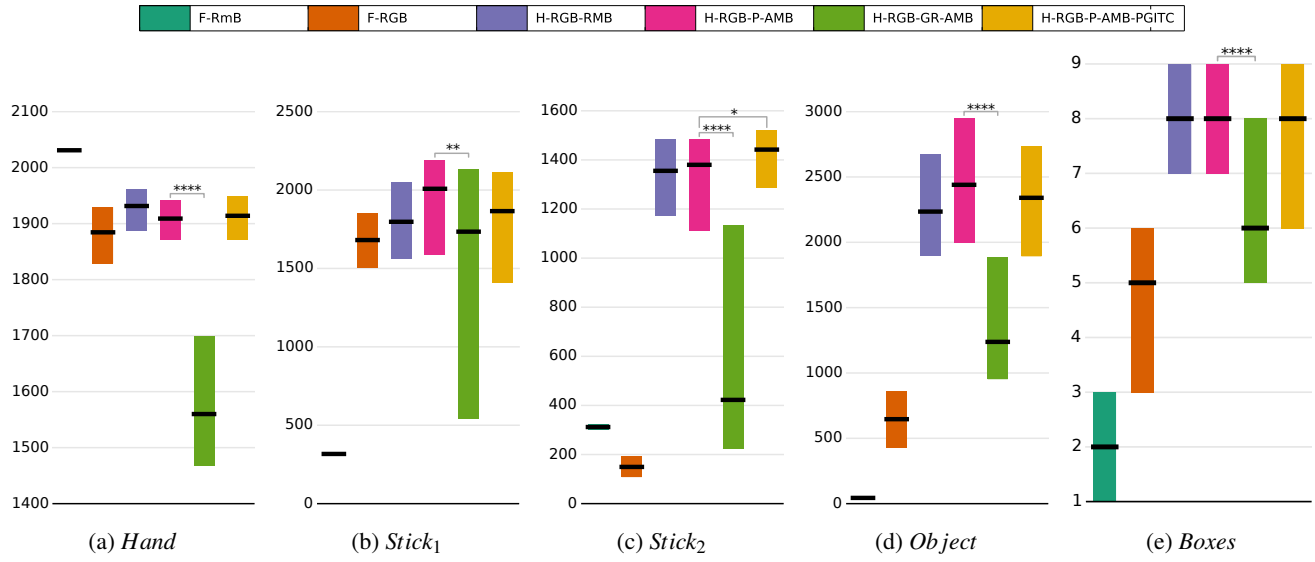


Figure 6: Exploration of sensory spaces.

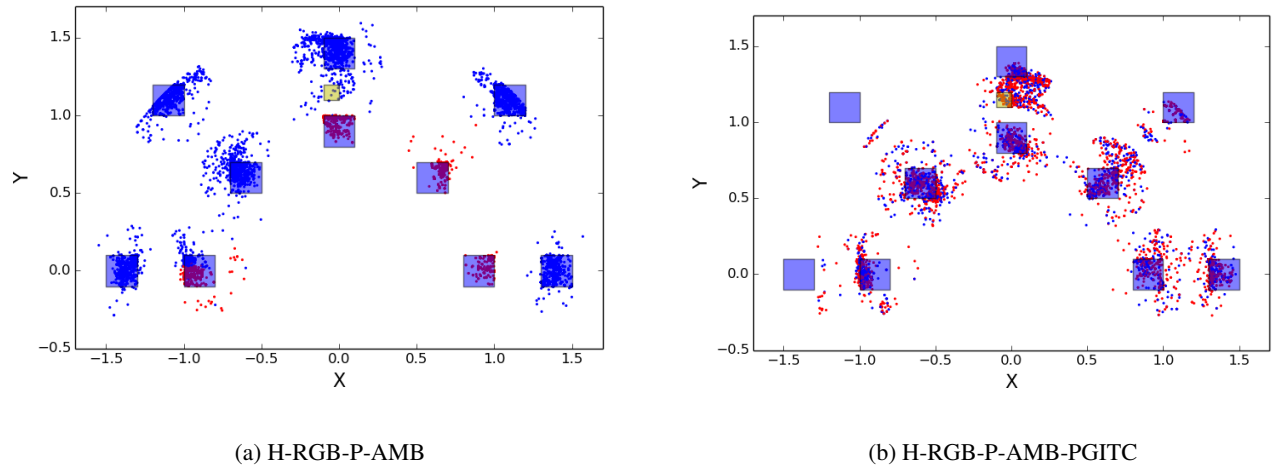


Figure 7: Chosen tool depending on object goal position. Blue points: long stick choice. Red points: small stick choice.