Acquisition and Utilization of Mental Imagery Capability in Robotic Action Sequencing Tasks

Kristsana Seepanomwan[1]

ABSTRACT

This work presents a series of neurorobotic models underlying learning in robots. It demonstrates the way in which, during sensorimotor exploration, robots do not just gain knowledge about how to form movement primitives but also obtain a mental imagery capability. Mental imagery plays a key role in these computational models by accelerating learning processes of action sequencing tasks. The first experiment involves permitting a humanoid robot to learn how to retrieve an out-of-reach object using a provided tool. This experiment explores a phenomenon on tool use development found in human infants. In addition, it tests two hypotheses on tool use development. The second experiment extends the domain of robot learning by targeting a useful robotic application. It drives a service robot to learn to acquire knowledge of how to manipulate perceived objects based on the objects' information and a goal from users. By means of planning, learning the sequence of actions in mind, the robots are able to learn by examining actions' outcome without really performing actions. This allows the second model to completely exclude parts of overt movements from the training loop. The results confirm that two types of robots can complete their given tasks in a reasonable period of time. The proposed models would benefit robotic applications in terms of online learning.

Keywords: Robot Learning, Action Sequencing, Mental Imagery

1. INTRODUCTION

Robot learning [1-3] has a goal to go beyond traditional preprogram techniques to control robots that lack adaptation. In this scheme, robots gather knowledge and identify users' goals by interacting with a specific environment. This kind of learning benefits robotic application in terms of flexibility, open-ended learning. A reinforcement learning framework is often used to underlie this success [3]. However, the nature of reinforcement learning requires a long period of exploration in order to fine-tune the underlie control system through trial and error. Unluckily, letting robots learn to work by themselves in the human environment cannot wait that long. This work demonstrates a way to overcome the issue of long exploration times. It presents two experiments that utilize the concept of affordances and mental imagery to shorter the learning period.

The first experiment introduces a neurorobotic model and learning scheme for robots and illustrates the developmental characteristic of how to achieve a given tool use task. The experiment focuses on replicating the processes that might be involved in the acquisition of knowledge about how to use tools in humans. This experiment is inspired by the idea that robots could do more useful tasks for us if they could utilize tools as we do. Through the assistance of affordances and mental imagery, the robot will gradually obtain tool use ability. Furthermore, the model will most likely reproduce an essential quality of tool use improvement found in human infants [4]. In infants, the advancement of how to use tools can be portrayed by the number of motor skills they have gained. Young infants ought to have fewer motor abilities because of the short timeframe (their age) they had in the sensorimotor learning stage. Subsequently, their tool use execution should be very limited. Older infants played out the tool use task better since they have procured an increasing number of motor abilities. The reason why having more motor abilities results in better execution is that it gives more chances that appropriate motor abilities for tackling a tool use scenario have been learned. In this setting, if there is no appropriate motor ability, the skills of using tools will not be conceivable. In addition, the experiment presents another speculation that different timeframes (preparation period) the infants used to familiarize with tool use situations would result in better tool use execution. Similar to the case of motor abilities, young infants will have fewer preparation times while older infants will have more. In terms of advancement, expanding these values, i.e., a number of skills, the period of preparing preliminaries, should result in a better performance.

In the second experiment, we have set out to explore more results on utilizing mental imagery in robots in terms of robotic application. A simulation

Manuscript received on June 30, 2019 ; revised on September 4, 2019.

Final manuscript received on November 1, 2019.

[1]The author is with Department of Computing Faculty of Science, Silpakorn University, Thailand., E-mail: krist49ers@gmail.com

of a service robot will be used as a test platform. This robot is capable of manipulating an object with a great range of flexibility and is suitable for applications such as object manipulation and transportation. In this experiment, the robot will be assigned a task of object classification on top of a table rather than bringing an object into reach.

The two experiments utilize the concept of affordances and mental imagery to help learning in robots. Affordance [5] can be explained as an ability to interpret perceived stimuli and prepare suitable actions to respond to them. This action preparation is possible through experience during life and allows humans and other animals to correctly react to various situations in the environment that are always changeable. In robots, this capability could help them increase the success rate of task completion for users. On the other hand, mental imagery [6-7] refers roughly to the processes that resemble perceptual experience. For instance, mental imagery can be stimulated hearing, seeing or visualizing, without the occurrence of the appropriate stimuli. It might be said that mental imagery occurrs in the mind (brain) like the phrase "seeing in the mind's eyes". Humans and other animals often use mental imagery, especially images, to underpin their plans or decisions. According to the theory of child development [8], children can utilize mental images when they get into the sensorimotor stage 6 at the age of 18 months. Mental imagery may have a vital role in the development of tool use and other cognitive/motor skills.

Despite, literature regards tool use development in human infants shed light on the utilization of mental imagery, the work on cognitive robots recommended the vital role of affordances [9-10]. However, there are no computational models that capture these issues so far. The first experiment is the first endeavour to reveal the role of both mental imagery and affordances in tool use development. The second experiment is focusing on utilizing mental images. In this paper, the capacity for using the tool and classifying objects in robots is conceivable through affordances and mental imagery, while the developmental characteristic is compelled by a number of motor skills. Essentially, mental imagery can be utilized in an action sequencing process, replacing the need for overt motor executions. The following four sections give more explanation on background concepts, settings and important components used in the proposed models. Section 6 gives a conclusion and suggests a future work.

2. BACKGROUND

2.1 Tool use development

The ability to use tools clearly helps humans to live their lives easier and achieve specific tasks more efficiently and conveniently. Other animals also have been reported to use of some tools such as rocks and wooden sticks to help cracking nuts or bringing food [11]. However, only in humans is found a great range of tool utilization. Inventions have been found, thanks to the humans' brain. And in-line with other cognitive skills, tool use's performance seems to vary with the humans' age. Younger age limits the knowledge of how to utilize tools, and sometimes is not possible at all. For example, in some easy situations using tools, such as taking a faraway object to reach with a provided rake-like tool, can pose difficulty to all humans aged younger than 18 months. It has been reported that at that age, the performance of using tools remains invariant and contingent. Rat-Fischer and colleagues [4, 12] conducted a series of interesting experiments that clearly show the relationship between tool use performance and the infants' ages. Fig. 1 illustrates the mean percentage of success for each initial condition (c1-c5). The spatial gaps correspond to the infants' ages. Tool use understanding seems to be connected to the experience that each infant had obtained during life, for example dufing play. In their tool use tasks, difficulty occurs when the spatial gap between the tooltip and the target object, a toy, is increased. This spatial gap causes some infants to fail in the test. It might be because tool use understanding requires knowledge about object-object interaction or the ability to utilize mental images. In their additional experiment, the infants that previously failed in the test can spontaneously succeed when giving them a demonstration of how to use the tool. However, only the infants aged 18 months and older that can learn from the demonstration. Thus, observing demonstrations (i.e., [4]) of how to use a tool might substitute the missing information about the action-outcomes effects of manipulating objects using tools to the infants.

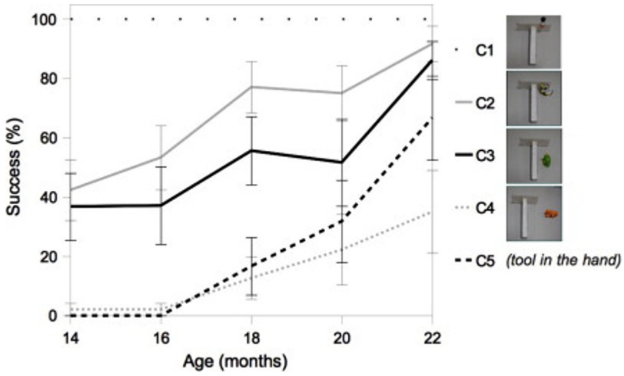


Fig.1: *The Development of Tool Use in Infants [4].*

On the other hand, tool use development in human infants can be viewed as highly related to their existing motor skills [13-15]. Due to their age, younger infants will have a smaller number of motor skills which directly limits their tool use performance. This is because suitable skills needed to complete the tasks have not been acquired by these infants. In contrast, obtaining more motor skills gives older infants more chances to succeed in the test. Some researchers argue that, in the beginning, infants can display only basic tool use ability using straightforward sensorimotor information alone [13], while in the later stage, the tool use utilization required the exact information about object-object interactions and the capacity to control an internal representation of that knowledge [16]. Affordances assume a focal role combining knowledge from perception and action [13]. At birth, human infants are not invested with a capacity to see affordance of artefacts. To comprehend their general surroundings, through affordances, they need to investigate this ability [17].

2.2 Cortical brain involved in the mental imagery

The Mental Rotation problem [18] is an excellent example case of mental imagery utilization. It involves comparing pictures or drawings of two objects that are arranged in different degrees and answering the question of whether the two objects are the same or the mirror version of each other. This problem has been studied and researchers have reported that the time spent in response and the error rates vary with the angular disparity between the two objects. An interesting question from this study asks why the answer to the imagery-based question seems to take time to respond as close to rotating objects in the real world. Considering that rotating mental images is not constrained by the laws of physics.

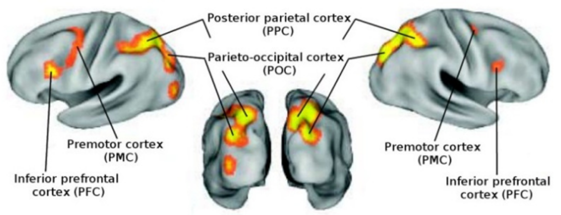


Fig.2: *The Cortical Brain involved in Mental Processes [19].*

In order to study the mechanisms of the brain related to solving this problem and how they work, most of the methods are scanning the human brain while someone is solving the Mental Rotation problem. It has been found that in addition to the Prefrontal Cortex, which is the part used to create mental images, the brain part that controls movements (Motor Areas), in the upper and back part of the brain, the Parietro-Occipital Cortex, also has significant activations (see Fig. 2) [19]. This suggest a relationship between motor activities and mental imagery. It has been believed that the two areas are connected because they are both used in sensorimotor learning. In humans, mental imagery is possible because they have rich experience of performing actions and observe the outcomes.

3. METHODOLOGY

3.1 Cognitive processing

A neurorobotic model presented in this study was constructed based on findings regarding the involvement of specific brain areas during the period of cognitive processing [19, 24]. The model also includes the intrinsic motivation mechanism that is believed to play a key role in shaping what agents can learn, and when they should start learning. The main components of the model are prefrontal cortex (PFC), parietal cortex (PC), primary visual cortex (V1), premotor cortex (PMC), primary motor cortex (M1), and the motivation activation mechanism including hippocampus (HIP) units.

The V1 is constructed as a two-dimensional neural map sized 320 × 240. This neural map is used to store camera images or visual information captured from the robot's eye. PC is also a neural map with the same size as V1. It encodes the spatial information extracted from the image in V1 using basic color filtering processes.

The process corresponding to skill acquisition of the model was underly by the motivation activation mechanism. This part was inspired by the idea of intrinsic and extrinsic motivation in psychology. Intrinsic motivation serves as an internal reward such as joy or surprise that the human infants receive when they have discovered some interesting events during play. Extrinsic motivation refers to the reward received from outside, for example food. In this work, we use the change of the environment, that is the current view of the users' task, as a source of attraction. The motivation activation works as a critic or dopamine neurons in the brain.

3.2 Learning mechanism

Information propagates through connections between neural maps. This study includes four connections and three types of learning algorithms: C1, C2, C3 and C4 (see Fig. 5, section 4). C1 and C2 are Hebbian connections that encode affordance. C3 uses supervised Kohonen training. This connection can spread the data between PFC_1 and PMC in both directions. Through this connection, a mental image can be created. C4 is a Q-learning connection. It accumulates knowledge of the action sequencing skills displayed during training.

Affordances represent the link between situations of tool use and their appropriate actions. The perceived situation activates an appropriate action that leads the robots to achieve a specific goal. It is formed during the acquisition of skills. To different robots, the same task, such as a tool use situation, can lead to the acquisition of different goals. The Hebbian learning rule is defined as equation 1:

$$\Delta w_{ij} = \eta \ a_i(a_j - w_{ij}) \qquad (1)$$

Equation 1 calculates the weight change Δw_{ij} using the activity level of post-synaptic neuron a_j subtracts by the current connection weight w_{ij} and multiplying the learning rate $\eta = 0.15$ and the activation of the pre-synaptic neuron a_i.

Mental imagery means the ability to forecast the outcome of the intended action. It will be used to support a robot's planning. Ideally, using mental images (in PFC_1) alone will be enough to support the acquisition of skills. The supervised Kohonen learning rule is given in equation 2.

$$\Delta w_{i\ win} = \eta \ a_{win}(a_i - w_{i\ win}) \qquad (2)$$

Equation 2 updates only the connection weights between one winning neuron in PMC to all neurons in PFC1($w_{i\ win}$). It uses the learning rate $\eta = 2.0$ multiply by the activation of the winning unit awin, then multiplied by the difference between activation of neurons in PFC1 a_i and the current connection weights $w_{i\ win}$.

Q-Learning is an innovative machine learning method mainly used in the study of action sequencing [25]. The Q-learning process accumulates knowledge for any given task based on action, state, and reward. The reward is used as a guide to select the next action and also affects the next state. The Q-Learning system can achieve an optimal solution via the rewarding and updating process (reward maximization). The Q-learning processes used here were adapted to train the connection C4 of the model as shown in equation 3.

$$\Delta w_{i\ win} = \eta((rewaed_t - \gamma \ qMax_t) - Qwin_{t-1})x_{i\ t-1} \qquad (3)$$

$reward_t$ refers to a value of 0 or 1 that will be assigned at the time step t. It will be subtracted by the product of the discount factor γ=0.8 and the maximum value of the q table $qMax_t$. After that, the maximum value of the q table in the previous time step $Qwin_{t-1}$ is subtracted. Finally, the multiplication with the learning rate $\eta = 0.0001$, and activation of the input units of the previous time step $x_{i\ t-1}$ will be performed.

3.3 Interesting events

The term intrinsically motivated events (IMEs) [20] refers to the events that distract the infants' attention. In this work, we define manually that what the robot can detect will indirectly refer to what skill the robot can discover. This work assumes that the IMEs are a source of motivation that drive the infants to practice the underlying action that causes them to develop skills. The individual difference is set according to the number of IMEs the robot can detect. Any interesting event will be revealed by chance, thus different robots can discover different events and acquire different skills. Furthermore, some skills may useful for solving a given task while the others may not be. Consequently, it is possible that different infants from the same age group can or cannot solve the same task.

Actions can be distinguished as intrinsically or extrinsically motivated by considering the intention behind them. For example, pulling action is an extrinsically motivated action if the intention of doing pulling is to bring food back for consuming such as when the infants are hungry. In contrast, if the intention is to bring other objects such as toys back for play, this might be the case of intrinsically motivated action. Therefore, pulling and interaction actions of this work will be determined as intrinsically motivated actions. These actions do not cause any external reward to the infants. Instead, they cause something interesting that distract the infants' attention such as the toy moving when touched with the tool. We assume that these kinds of events make the infants keep doing those actions in order to constantly make interesting events happen.

In order to demonstrate that robots can acquire the knowledge of how to make useful actions in a reasonable amount of time, we have established a scheme where the change of the scene caused by movements of the robots is appropriate for a simple scenario for the use of the tool and object classification tasks. For example, six types of situations are assumed to occur during exploration from this interaction (movement of the robot's arm with the tool, see section 3 for the details). In infants, when these events happen, they may be surprised or distracted because the object of desire (a toy) was moved. Interesting events include pulling, touching and moving the toy in four directions: North, South, East, and West.

3.4 Dynamic Movement Primitive

DMP's framework [21] was used to create a set of motor skills for robots. Encoding of motor skill or movement trajectories is carried out by a number of linear and non-linear dynamic functions. DMP can generate a series of joint angles that when rolled out, can be directly used to control robotic control actuators. Equations 4, 5, and 6 are key equations of this framework.

$$y_t^{''} = \alpha_y(\beta_y(g - y_t) - y_t^{'}) + f_t \qquad (4)$$

$$\psi_i = \exp(-h_i(x - c_i)^2) \qquad (5)$$

$$x^{'} = -\alpha_x x \qquad (6)$$

The first part of equation 4 is a linear dynamic system that is shaped by the forcing term f. The trajectory y_t will be used to control one specific joint of the robot. The forcing term f is calculated by normalizing the weight values (Eq. 5) with the canonical value x which was reduced by the factor α over time, as specified in Eq. 6. The weight value ψ_i has been defined as a Gaussian function with variance h, center c, and activation x.

Unlike a traditional use of the DMP, there are no observed movements provided in this work. The system must find the right move through the processes of trial and error. This was inspired by the way infants acquired their motor skills. This work applied the Policy Improvement Black Box optimization (PI^{BB}) [22] algorithm to learn the movement action primitives. Using PIBB does not require any target trajectory for comparison. Instead, it determines the rewards produced by several samples. Initially, the PI^{BB} creates some sample movements by adding random noises to the current parameter sets (shapes, goals) of a DMP. A reward assigned to each sample can be calculated arbitrarily depending on each tasks' specification. Ideally, each sample will get a different reward. The new parameter sets of the actual control will be calculated by averaging the parameters sets of all sample, weighted by their rewards.

The PI^{BB} algorithm consists of 4 main steps: adding noise, awarding, weighting, and averaging. These will be repeated until an outcome is satisfied or reach a maximum step is reached. Initially, the goal parameters of the DMP (joints' angle) will be assigned using the value of a given smart move, while the shape parameters will be calculated using weight averaging technique borrowed from the basic DMP. These initial values will be called the mean parameter set. After that, all of these parameters will be adding up with noise terms which, when the DMP roll out, will cause different movements.

The main idea behind the PI^{BB} algorithm is that it finds a new set of a policy's parameters by averaging all created samples based on their rewards. By means of weight averaging, the new mean parameters will lead a system in the right direction, approaching a goal. This process will be repeated until the maximum cycle count is reached or an action based on the new mean parameters gains enough reward (cause an interesting event). The PI^{BB} algorithm is well suited to learn shape parameters in the DMP's framework. The DMP's framework was applied to create actions, and the PI^{BB} was used to refine them. Fig. 3. Illustrates the difference between basic DMP and DMP+PIBB. The left part of Fig. 3. illustrates that basic DMP learn to imitate an observed trajectory by minimising its cost. In contrast, on the right part, a goal direct DMP learns to acquire new motor skill by exploration.

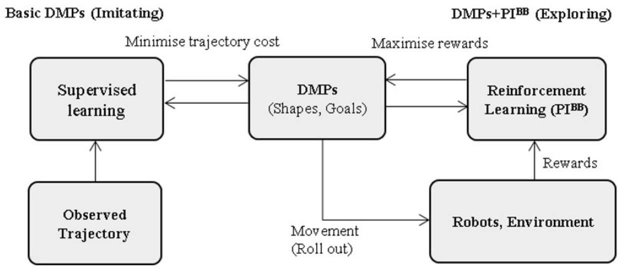


Fig.3: *The difference between basic and a reinforcement learning DMP.*

4. EXPERIMENT-1: TOOL USE DEVELOPMENT

4.1 Robotic tool use scenario

The iCub simulator [23] was used to simulate infant participants. In front of the robot is a table, a big rigid box (see Fig. 4a). Scenarios for the use of tools are set at the top of this table. At a location too far from the robot's reach, a toy, the red cylinder, will be placed on the table. The iCub was set to use its' right hand to hold a rake-like tool permanently. The tool is in green color with a long stick handle and a flat rectangle tip. The tool has been used to extend the robot's reach length. The competence to use the tool requires only the robot's right arm. Movements caused by the right arm directly affect the position of the tool.

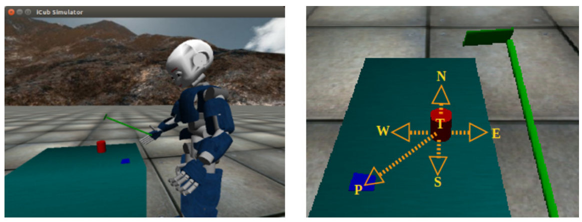


Fig.4: *Robotic Tool Use Scenario.*

4.2 Initial postures and tool use situations

In fact, it can be a daunting task to demonstrate the processes of acquiring all possible actions or even focusing on one arm moving with a tool. Therefore, for our initial tool use scenarios, only two types of actions that the robot displays with the tool are considered mandatory. They are actions of "pulling" and "interaction." What differentiates each action is its result. Pulling, for example, is an action that moves the toy to an accessible area, while the interaction actions refer to the effects when the tool interacts with the toy, for example touching it, or moving it to the left.

The term initial posture refers to the configuration of the robot's arm. Changing the value of the joints directly causes the posture of the robot to change.

There are four initial postures that vary the spatial gap between the tip of the tool and the toy. The first posture places the tip of the tool behind the toy. The second has a small spatial gap, while the third and fourth have greater gaps. In order to simplify the exploration of action-perception and the acquisition of tool use skills, we assumed that the robots hold the rake tool permanently in their right hand and will only encounter four different situations during the action-perception exploration period. From the point of view of the robot, its visual perception is directly affected by movements of its arm. This causes changes in the spatial difference between the tip of the tool and the toy.

4.3 Architecture of the model

The computational model of the first experiment (Fig. 5) includes several parts of the cortical area as stated in the methodology section. In this experiment, the robot was set to stand still in front of the table. In most cases, it is able to see both the tool and the toy. Tool use situation will be taken as an image from its right eye and will be sent to the model. By using a basic color filtering process, only parts of the tool and the toy will be retained and used as neural activation of the map V1. Thus, changing the perceived scene will cause a change in the neural activation. Different tool use scenarios are variations caused by movements of the robot's right arm and the interaction between the tool and the toy. Some movements could cause the toy to move in a certain direction when touching the tool, which are considered interesting situations.

There are 10 neurons in the PMC map. They are used as affordance interpretation. Activations of these neurons refer to the preparation of action. Each neuron in this map has a direct link to a specific unit in M1. What is used to construct units in M1 is not a neural mechanism but a form of motor execution. Thus, activation of the PMC's neurons causes execution of the corresponding M1's unit. There are two neural maps in the part of PFC: PFC_1 and PFC_2. The PFC_1 is used to encode mental images created by the connection C3. PFC_2 is used to store the goal. There are 6 neurons in the HIP. Each neuron is used to monitor the specific interesting situation that will be interpreted by the connection C1.

4.4 Action acquisition

The DMPs were used to control movements of the right are of the iCub simulator. The DMPs have been designed to have 7 goal parameters that match the number of joints, and 28 shape parameters (4 for every joint) to modify the trajectory of the movement. The two sets of parameters can be automatically controlled for each joint of the arm.

In order to obtain proper parameters for the DMPs, a form of normalization is applied. We have

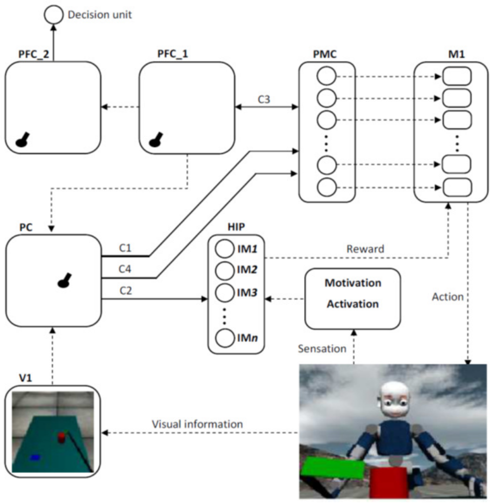


Fig. 5: *The Neurorobotic Model for Learning to use Tool.*

specified that the model starts with random shape and goal parameters. Thus, the robot begins the sensorimotor learning by performing random movements. As soon as it has discovered new events (no activation on the HIP's neurons), the goal parameters of the current DMP will be used for the other 10 DMPs. Therefor the normalization is done over 10 sets of shapes parameters from 10 DMPs (samples). Therefore, a set of normalized goal and shape parameters will correspond to one specific skill or a movement that causes an interesting event.

The number of interesting events is restricted. Each simulated infant will be able to discover 6 interesting events. However, in fact, there are some cases where the model uses more than one neuron to achieve the same goal. The individual difference means each robot has acquired different DMP parameters and attached them to different PMC neurons.

4.5 Action sequencing

In the Q-Learning process, the action is determined by a winning neuron in the PMC which is directly connected to a specific DMP. States refer to the activation of the PC. The reward of 1 will be received when the robot can bring the toy to the target position.

Fig. 6 shows the strength of connection weights (C4) between 4 different initial tool use situations and the PMC's neurons. As stated before, the two examples (Fig. 6a and 6b) correspond to the two individuals. Four images displayed in PC are used for illustration purposes. In fact, the PC encodes only the part of the tool and the toy.

The difference between the two examples is determined by the variety of PC-PMC connections. Furthermore, PMC activation should differ as an affordance interpretation, but this is not displayed in the

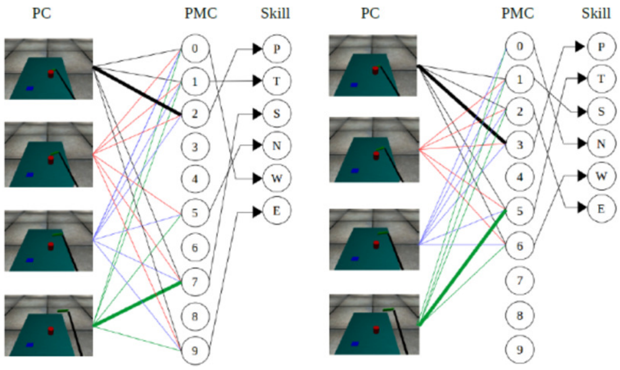


Fig.6: *Examples of Individuals Difference as Connection weight strengthen.*

figure. The thicker lines refer to the Q-learning connections, which are strengthened corresponding to the initial position encoded in the PC. Lines with different colors are used to illustrate different initial postures.

In example 1 (Fig. 6a), the PMC neuron number 2 is activated with high activation when the robot encounters the initial tool use situation number 1. This means that, because of the SoftMax function, the robot is more likely to choose Action P. Likewise, the PMC neuron number 7 is activated with the highest activation when the robot sees tool use situation number 4. The system is expected to exhibit action S. Given that the outcome of action-S is similar to the situation of the tool use number 1, the system should then activate action-P and this is inturn should lead to the success of the toy being retrieved.

In contrast, if the system is formed as in example 2 (Fig. 6b), it cannot succeed in the case of a tool far from a toy (tool use situation number 4). The reason for this is that the PMC neuron number 5 in the initial tool use situation number 4 is active, but it was mapped to the action-T that is usually not like the situation number 1 of tool use. This system is therefore supposed to fail in the test.

4.6 Experiment settings

As revealed by [4], the performance of the use of tools is limited by the age of the child. Increasing age leads to the better performance in the test. We have interpreted the age as number of motor skills and training trials and conduct two tests based on two hypotheses:

- Hypothesis 1: In this test, the age was attached to the different number of motor skills. This models the fact that younger infants should have a small number of things they can do, while older infants can do many more things because they have acquired a higher number of motor skills. The approach is to restrict the case of interesting events that each simulated infant can detect. The youngest infants can detect the fewest cases of interesting events and the number of events detected gradually increased as they grew up. Thus, infants aged 14 to 22 months (i.e. 14, 16, 18, 20 and 22) were created by varying the number of the interesting events they can detect from 2 to 6. Note that, all robots in the simulation were initialized with two fundamental motor skills: pulling and touching, and they all use the same training cycle count of 10.
- Hypothesis 2: In this test, the number of training trials is used as the infants' age. All simulated infants will be equipped with all 6 motor skills, but the number of training trials used in the Q-Learning processes will vary. Younger infants are simulated using a small amount of training, while the older ones will be simulated using a higher amount of training. In detail, the simulated infants aged 14 months refers to the training cycle count of 2, while robots aged 16, 18, 20 and 22 months are described by the training cycle counts of 4, 6, 8 and 10.

In each test, the simulated infants were required to solve a given tool use task in one of two situations: "tool behind toy" or "tool far from toy". In addition, one of the two solving strategies must be utilized by each robot: "Reactive" or "Planning". The "Reactive" strategy refers to the case when the robot exhibits overt movements during learning, while "Planning" is the case when the robot uses mental images instead. By means of a statistical test, each experiment will be repeated 30 times. Thus, we needed 60 individuals.

4.7 Results

The first experimental results are illustrated in Fig. 7. The top line (blue) of the two diagrams (Fig. 7a and 7b) shows the performance of the use of the tool in the situation "tool behind toy". The performance of all age groups' tools is similar in this case, with a high success rate of around 90%. The youngest group seems to be able to recover the playground at the highest success rate, 100%. This is because they have only two skills in their actions, i.e., pulling and touching, and the situation of the "tool behind toy" is more likely to lead to high activation on the skill-P due to the affordability interpretation. Thus, the first group gains benefit and this leaves no room for improvement. Increasing skills will therefore not lead to performance improvement.

The lower line (red) of the two graphs in Fig. 7, in contrast, shows a progression in tool use performance when more skills are available. Unlike the case of "tool behind toy", skill-P cannot be used to retrieve the toy from the situation where it was set to have large spatial gap ("tool far from toy"). Therefore, the performance of this begins in the smallest group from approximately 0% and increases to 10, 40, 45 and 60 corresponding to the increase in the infant's age.

Fig. 8 illustrates the second experiment's results.

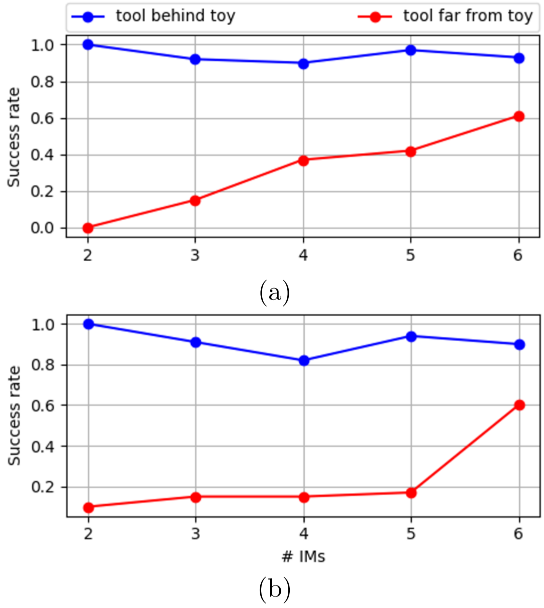


Fig. 7: The Result of Simulating the Infant's Age as the Number of acquired skills.

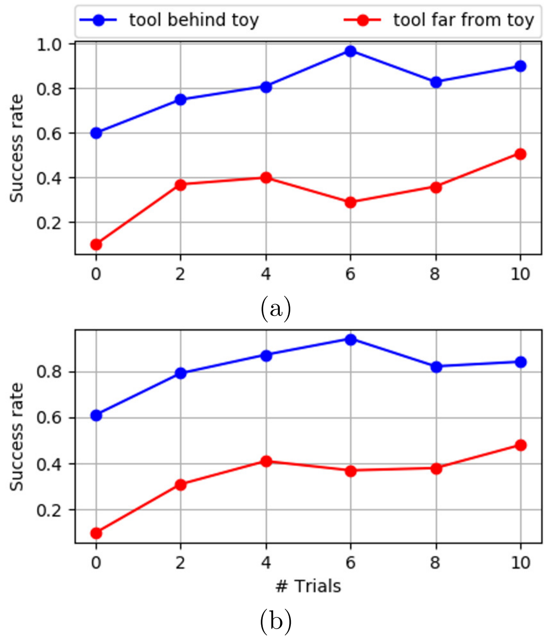


Fig. 8: The Result of Simulating the Infant's Age as the Number of Training Trials.

The youngest group without training can resolve the tool-behind-toy case at approximately 60% of the time in this test. Unlike the first experiment, roughly 20% of the individuals in the group can solve the case of "tool far from toy". While the increase in the number of training trials leads to an increase in the performance of the two tests, the performance feature is not linear. It seems that the use of the amount of training cycle counts of 6 results in the highest performance in the case of "tool behind toy".

The results obtained by the two different strategies, "reactive" and "planning", appear identical in both tests. This confirms that the use of open movements can be effectively replaced by mental imagery. However, the first hypothesis seems to capture the main feature of the development of tool use better than the second when compared to the children's data reported in [4].

5. EXPERIMENT-2: OBJECT CLASSIFICATION

5.1 Robotic Application

In this new experiment, a study on robotic manipulation is introduced. It focuses on object classification based on color. A simulation of Fetch robot provided by Fetch Robotics (`https://fetchrobotics.com`) was used as a test platform. This robotic simulation supports the Robot Operating System (ROS) framework (`https://www.ros.org`), possible to integrate it into a Gazebo simulation (`http://gazebosim.org`). Finally, it is more suitable for robotic application such as object manipulation, transportation than the iCub. A Fetch robot is a mobile robot that has a mobile base, an articulated arm with a gripper, a torso, and a pan-tilt head with a depth camera. Both physical and simulated robots have been widely used as test beds in robotic research in both the educational and industrial domains. Fig. 9a illustrates the simulation of the Fetch robot. Fig. 9b is the image of a red object on the table from the robot point of view. Fig. 9c illustrates that when performing action, the object will be occluded by part of the robot arm.

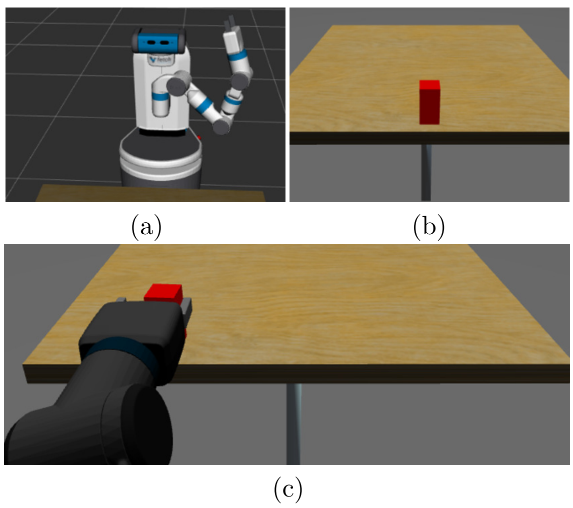
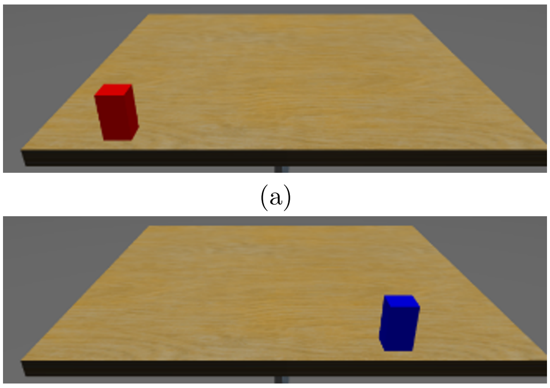


Fig. 9: Simulation: a) The Fetch robot b) an Initial Task c) The Robot's Arm in Action.

The robot will be facing situations where there is an object placed on top of table. Its goal is to move the object, with its gripper, from various initial positions to some specific target locations on the table defined by users. In short, the robot has to discover the users' goal and perform a sequence of actions in order to reach the goal. For example, when seeing a "red" object, the goal of moving it to the left may be assigned by users (Fig. 10a). On the other hand, if the "blue" object is placed on the table, the goal might be moving it to the right (Fig. 10b). In order to complete this task, the robot must discover effect on action. This task clearly requires a set of actions to accumulate knowledge of action sequencing. By means of reinforcement learning, goals are targets that an agent tries to accomplish. In this test, the goal is to find and move an object to a specific target location.



(a)

(b)

Fig.10: Samples of User Defined Goals: a) red object on the left b) blue object on the right.

5.2 Architecture of The Model

Architecture of the model of the second experiment is illustrated in Fig. 11. It is a modified version of the one detailed in the first experiment. However, motivation activation and HIP are excluded since we have done the action acquisition processes offline to speed up the experiment. The connection C1 is also an affordance connection as in the first experiment. However, the connection C2 is a Kohonen connection that is trained to create mental images. The last connection, C3, is a Q-learning connection. Each connection has been trained using the processes as described in section 3.3.

The new model is also equipped with the parts of beliefs and object information which play an important role in the training of the connection C1 and C3. These additional signals will help the model to differentiate the case when no part of the robot's arm occluded the object. The map PMC has 8 neurons corresponding to 8 actions. The belief unit also has 8 neurons that encode a state of the robot. Color detection is used to identify color of the object of a given task and will be stored in the object information units. One constraint for this experiment is that there will be only one object with one color on the table.

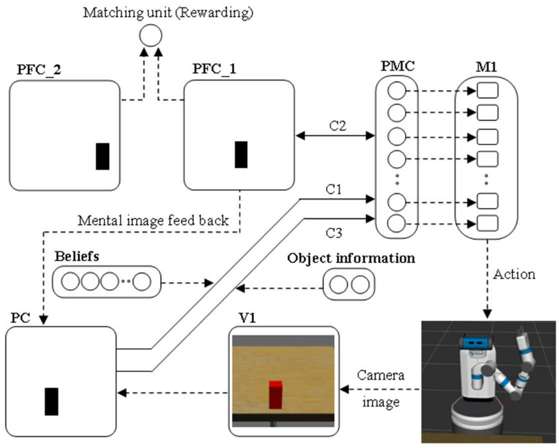


Fig.11: The Neurorobotic Model for Object Classification.

5.3 Actions' Definition

This experiment defines actions as movements of the arm to accomplish some purpose. The arm of the Fetch robot has 7 joints with a great range of movement allowing it to do complex movements. It is important to note that, unlike the iCub, joints of the Fetch are not constrained by the human arms, thus it can reach a goal by exhibiting weird movements. In this experiment, MoveIt Planning Framework [26] is used as a controller to control the robot arm instead of the DMP framework. Its benefit is providing a set of great tools such as inverse kinematic control and obstacle avoidance capabilities. Thus, action acquisition processes are constructed more easily, more quickly, and are more flexible without the need for sample movements, compared to using DMPs as in the previous experiment.

In order to tackle the task of manipulating objects and learning to classify them by colors, 8 basic actions are defined. Each action is a movement of the arm that brings the robot's gripper to a specific location in space with respect to the object. This can be done by finding the location of the object, [x, y, z], and giving it as an input to the MoveIt controller. See table 1 for the details. Action ML and MR do not update the x and z values since the two variables can be taken from the previous action. However, they are the only actions that can lead the model to achieve the users' goals. In detail, x, y, z are the location of the object on the table relative to the position of the robot.

From the first experiment, we have realized that, in order to utilize the mental image in planning, the robots must have a complete set of actions or all mandatory actions. Planning based on a limited number of mandatory actions will not lead to a good performance because there is no chance that correct combinations of how to achieve a given task can be found. Thus, for the sake of pursuing robotic applications, the robots will able to perform all basic actions stated in table 1.

Table 1: *Basic Actions.*

No.	Name	Description	Config.
1	PC	**P**repare the gripper on the **C**enter poses with respect to the object location	x-0.3 z+0.1
2	PL	**P**repare the gripper on the **L**eft of the object	x-0.3 y+0.1 z+0.1
3	PR	**P**repare the gripper on the **R**ight of the object	x-0.3 y-0.1 z+0.1
4	AC	**A**pproach the object from the **C**enter	x-0.15 z+0.1
5	AL	**A**pproach the object from the **L**eft	x-0.15 y+0.1 z+0.1
6	AR	**A**pproach the object from the **R**ight	x-0.15 y-0.1 z+0.1
7	ML	**M**ove object to the **L**eft	y=0.3
8	MR	**M**ove object to the **R**ight	y=-0.3

The states or beliefs of the robot (Table 2) are constructed based on its last action. For instance, if the robot performs action "AC" its state will be "AC" which means that it believes it is in the state "AC". This experiment uses 8 states as shown in Table 2. This information will help the robot to distinguish each perceived scene and drive the model to select correct action. Ideally, each state will be used as an additional signal to train the connection C1.

Table 2: *States/Beliefs of the Robot.*

No.	Name	Description
1	Init	The robot is in the Initial state with the arm in the stow pose (see Fig. 9a)
2	PC	The robot is in the state of Preparing the arm at the Cente of the object
3	PL	The robot is in the state of Preparing the arm at the Left of the object
4	PR	The robot is in the state of Preparing the arm at the Right of the object
5	AC	The robot is in the state of Approaching the object from Center
6	AL	The robot is in the state of Approaching the object from Left
7	AR	The robot is in the state of Approaching the object from Right
8	N/A	This state is assigned when the robot performs action ML or MR. This state will not affect the connection c1

5.4 Experiment Settings

In the first experiment with the iCub's tool use development, the constraints that affect the performance were the number of motor skills and training cycles used in the Q-learning processes. However, as seen in the first experiment, varying the number of training trials did not have much affect on the performance, so the training cycles in this test is set to 10. In addition, since this experiment concerns testing the possibility of using mental imagery in robot learning, training the model using overt movements will be excluded. It tests the model with different initial positions and colors of the object that will be considered as an initial task for the robots.

Individual difference is created by different connection weights of the model that directly control movements of the robot. We create new simulated robots for the test by assigning new sets of connection weights to them. By re-training the connection weights (C1, C2, C3), two robots will react to the same task differently. In term of affordance interpretation, we assume that different robots will have different experiences with objects during the exploration phase. Thus, the level of the PMC's activation of each robot in response to the same perceived object will probably not be the same.

5.5 Results

By testing the model with 100 simulated robots as described above, we found that only the robot with good affordance interpretation (about 10%) will result in a good performance completing the task. Improper interpretation (about 90%) will not allow the robot to achieve all of the users' goals. However, since the connection C1 is trained on a scheme of sensorimotor learning, the result is unpredictable. In order to make the model respond properly, we must repeat the re-training process until the proper response is found. Please note that each robot was trained only on the cases that the "red" or "blue" object was placed on the "center" of the table. In testing, the objects were placed randomly. Fig. 12 shows samples of PMC's activation of different robots in response to the same initial task. The bar chart of Fig. 12a illustrates the case of a proper PMC's activation. Different activation of neuron PC, PL, PR will benefit the starting of the Q-learning processes. On the other hand, similar levels of activation of the three neurons as shown in Fig. 12b made the model unable to learn the object classification task correctly.

When the PMCs activation had the pattern as of Fig. 12a, it was possible of the robot to complete 100% of the 10 tests. In contrast, the pattern in Fig. 12b made it possible to only complete half of the test, or for only on one color, thus the maximum success rate is 50

6. CONCLUSION

The first experiment demonstrated and confirmed that the development of how to use tool can be progress with age. The first test (Hypothesis 1) interprets infants' age as the number of acquired motor skills, whereas the latter (Hypothesis 2) differentiates

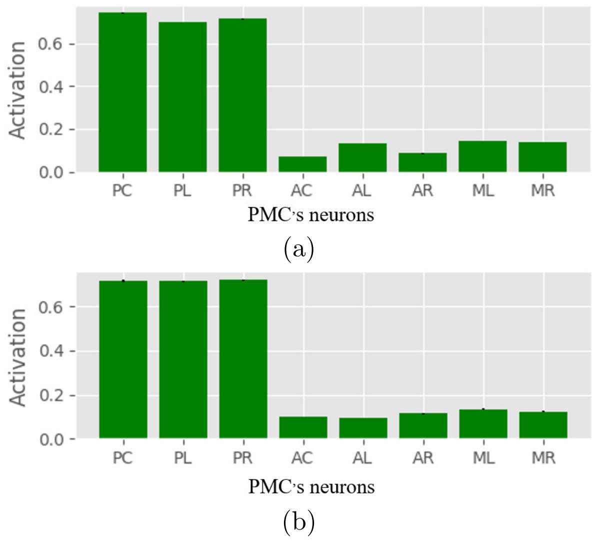


Fig.12: *Samples PMCs Activations a) A Proper Activations b) An Improper Activation.*

the age as a period of tool use experience before testing. Even though the results showed a similar characteristic for tool use performance (Fig. 7, 8), we suggest that the development of tool use found in human infants might be caused by differences in the number of motor skills they have obtained. However, there is still no evidence to support the connection between the number of motor skills or the experience period to the age. The performance of the use of instruments in human infants remains unknown.

Scenarios of tool use captured from the eye of the robot are changed subject to movements of the right arm of the robot that holds the tool. However, the number of different tool use situations does not simply mean the number of robot actions (2 to 6). Due to the unpredictable effect of the physical interactions between the tool and the toy, the Q-Learning processes forms knowledge of how to use tool from a variation of inputs rather than a static set.

This work also shows that mental imagery (an anticipation of the results from intended actions) can be used to replace overt movements for the achievement of tool use capability. Since the robots can plan to solve the task using mental images, this may be interpreted as an understanding of how to use a tool. In such cases, the robots can spontaneously use motions to solve a perceived tool use scenario like what happened in real human infants.

The second experiment confirms another possibility of using mental images in learning action sequencing. The concept behind this success is that the model can utilize mental images alone in training of the Q-learning connection. It is fast because there is no need to do overt movements to explore the given goals. In terms of robotic applications, this capability would help robots to learn the user tasks on the fly.

The two models use static images and compare their difference in order to identify the reward. In future work we will use camera images directly for the activation of the PC by utilizing a convolutional neural network (CNN) framework. This addition would allow the future model to deal with variance of object location and extend the work to other types of robotic applications.

References

[1] J. Connell and S. Mahadevan, Robot learning, vol. 233, Springer Science & Business Media, 2012.

[2] J. Kober, J. Bagnell and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32(11), pp.1238-1274, 2013.

[3] B. Argall, S. Chernova, M. Veloso and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57(5), pp.469-483, 2009.

[4] L. Rat-Fischer, J. K. O'Regan and J. Fagard, "The emergence of tool use during the second year of life," *Journal of experimental child psychology*, vol. 113(3), pp.440-446, 2012.

[5] J. Gibson, The ecological approach to visual perception: classic edition, Psychology Press, 2014.

[6] S.M. Kosslyn, Image and the Brain. The Resolution of the Imagery Debate, MIT Press, A Bradford Book, Cambridge, 1994.

[7] N. Thomas, "Mental Imagery," The Stanford Encyclopedia of Philosophy (Spring 2018 Edition), Edward N. Zalta (ed.), [Online] Available: `https://plato.stanford.edu/archives/spr2018/entries/mental-imagery/`

[8] J. Piaget, and M. Cook, The origins of intelligence in children, New York: International Universities Press, vol. 8(5), 1952.

[9] A. Stoytchev, "Behavior-grounded representation of tool affordances," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, (April. 2005), pp.3060-3065, 2005.

[10] V. Tikhanoff, U. Pattacini, L. Natale and G. Metta, "Exploring affordances and tool use on the iCub," in *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*, (October. 2013), pp.130-137, 2013.

[11] V. Bentley-Condit, "Animal tool use: current definitions and an updated comprehensive catalog," *Behaviour*, vol. 147(2), pp.185-32A, 2010.

[12] J. Fagard, L. Rat-Fischer and K. O'Regan. "The emergence of use of a rake-like tool: a longitudinal study in human infants," *Frontiers in psychology*, vol. 5, pp. 491, 2014.

[13] J. Lockman, "A perceptionaction perspective on tool use development," *Child development*, vol. 71(1), pp.137-144, 2000.

[14] B. Kahrs and J. Lockman, "Building tool use from object manipulation: A perceptionaction perspective," *Ecological Psychology*, vol. 26(1-2), pp.88-97, 2014.

[15] E. Gibson and A. Pick, An ecological approach to perceptual learning and development, Oxford University Press, USA. 2000.

[16] F. Guerin, N. Kruger and D. Kraft, "A survey of the ontogeny of tool use: from sensorimotor experience to planning," *IEEE Transactions on Autonomous Mental Development*, vol. 5(1), pp.18-45, 2013.

[17] E. Gibson, "Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge," *Annual review of psychology*, vol. 39(1), pp.1-42, 1988.

[18] R. Shepard and J. Metzler, "Mental rotation of three dimensional objects," *Science*, vol. 171(972):701-3, 1971.

[19] J. Zacks, "Neuroimaging studies of mental rotation: a meta-analysis and review," *Journal of cognitive neuroscience*, vol. 20(1), pp.1-19, 2008.

[20] P. Oudeyer and F. Kaplan, "What is intrinsic motivation? A typology of computational approaches," *Frontiers in neurorobotics*, vol. 1, pp.6, 2009.

[21] S. Schaal, "Dynamic movement primitives-a framework for motor control in humans and humanoid robotics," *Adaptive motion of animals and machines*, Springer Tokyo, pp.261-280, 2006.

[22] F. Stulp and O. Sigaud. "Policy improvement methods: Between black-box optimization and episodic reinforcement learning," 2012.

[23] V. Tikhanoff, A. Cangelosi, P. Fitzpatrick, G. Metta, L. Natale and F. Nori, "An open-source simulator for cognitive robotics research: the prototype of the iCub humanoid robot simulator," in *Proceedings of the 8th workshop on performance metrics for intelligent systems*, (August. 2008), pp.57-61, 2008.

[24] K. Seepanomwan, D. Caligiore, A. Cangelosi and G. Baldassarre, "Generalisation, decision making, and embodiment effects in mental rotation: a neurorobotic architecture tested with a humanoid robot," *Neural Networks*, vol. 72, pp.31-47, 2015.

[25] C. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8(3-4), pp.279-292, 1992.

[26] I.. Sucan and S. Chitta, "MoveIt", [Online] Available: `http://moveit.ros.org`

Kristsana Seepanomwan is a lecturer at the Department of Computing, Faculty of Science, Silpakorn University, Thailand. In 2007 he obtained an MSc (Computer Science) from King Mongkuts' University of Technology, North Bangkok (KMUTNB), Thailand. In 2011 he got a scholarship from the Thai Government for doctoral program in the field of robotics and enrolled as a PhD candidate at the Centre for Robotics and Neural System (CRNS), School of Computing, University of Plymouth, UK. In 2016 he finished the PhD (Computing) with thesis title "Mental Imagery in Humanoid Robots". His research interests concern cognitive robotics, robot learning, and brain-like processing mechanisms, with a focus on decision making, mental imagery, and planning in humanoid robots.