

The “Mechatronic Board”: A Tool to Study Intrinsic Motivations in Humans, Monkeys, and Humanoid Robots

Fabrizio Taffoni¹, Domenico Formica¹, Giuseppina Schiavone¹, Maria Scordia¹,
Alessandra Tomassetti², Eugenia Polizzi di Sorrentino³, Gloria Sabbatini³,
Valentina Truppa³, Francesco Mannella⁴, Vincenzo Fiore⁴, Marco Mirolli⁴,
Gianluca Baldassarre⁴, Elisabetta Visalberghi³, Flavio Keller², Eugenio
Guglielmelli¹

¹ Laboratory of Biomedical Robotics and Biomicrosystems,
Università Campus Biomedico, Rome, Italy
{f.taffoni, d.formica, g.schiavone, m.scordia,
e.guglielmelli}@unicampus.it,

² Laboratory of Developmental Neuroscience,
Università Campus Biomedico, Rome, Italy
{a.tomassetti, f.keller}@unicampus.it

³ Unit of Cognitive Primatology,
Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy
{eugenia.polizzi, gloria.sabbatini, valentina.truppa,
elisabetta.visalberghi}@istc.cnr.it

⁴ Laboratory of Computational Embodied Neuroscience,
Institute of Cognitive Sciences and Technologies, CNR
{francesco.mannella, vincenzo.fiore, marco.mirolli,
gianluca.baldassarre}@istc.cnr.it

Abstract. In this chapter the design and fabrication of a new mechatronic platform (called “Mechatronic Board”) for behavioral analysis of children, non-human primates, and robots are presented and discussed. The platform is the result of a multidisciplinary design approach which merges indications coming from neuroscientists, psychologists, primatologists, roboticists, and bioengineers, with the main goal of studying learning mechanisms driven by intrinsic motivations and curiosity. This chapter firstly introduces the main requirements of the platform, coming from the different needs of the experiments involving the different types of participants. Then, it provides a detailed analysis of the main features of the Mechatronic Board, focusing on its key aspects which allow the study of intrinsically motivated learning in children and non-human primates. Finally, it shows some preliminary results on curiosity-driven learning coming from pilot experiments involving children, capuchin monkeys, and a computational model of the behavior of these organisms tested with a humanoid robot (the iCub robot). These experiments investigate the capacity of children, capuchin monkeys, and a computational model implemented on the iCub robot to learn action-outcome contingencies on the basis of intrinsic motivations.

1 Introduction

Behavioral sciences encompass all the disciplines that explore the activities and the interactions among organisms in the natural world. They involve a systematic rigorous analysis of human and animal behavior through controlled experiments and naturalistic observations (Klemke, 1980). Behavior is anything that a person or an animal does. In behavioral studies, several approaches to observe and measure human and animal behaviors are used (Martin and Bateson, 1998). Especially in the past, while psychologists focused on the proximate causation of behavior and on general processes of learning in few animal species (namely those that better adapted to laboratory conditions), ethologists were typically interested in studying the ultimate causation of behavior, especially in nature where spontaneous behaviors and the role played by the environment could be better appreciated. These two fields are now increasingly integrating (Wasserman and Zentall, 2006), also thanks to neuroscience that contributes to clarify the neural mechanisms underlying observable behaviors.

The autonomous acquisition of new skills and knowledge is one of the most astonishing capacities that can be observed in humans and animal models. The driving force that shapes this process is unknown. Children seem to acquire new skills and know-how in a continuous and open-ended manner (Kaplan and Oudeyer, 2007). Before developing tool-use ability, for example, children show typical exploratory behaviors based on trial and error which could be considered as a self-generated opportunity for perceptual learning (Lockmann, 2000). Most importantly, this process is not directly related to biologically relevant goals but seems to serve the acquisition of skills and knowledge themselves. According to Thelen and Smith (1994), this process follows a well-defined path strictly linked to the development of cognitive and morphological structures, which are related to the new acquired skills (e.g. tool use). How children learn to use these skills in a different context to reach a specific goal is unknown. Also non-human primates show the capability of learning new skills through exploratory behaviors (see White, 1959). Even if distinguishing food-seeking exploration from generic exploration is problematic, several studies have shown that nonhuman primates learn to efficiently manipulate mechanical puzzles whose solution is not rewarded with food or water (Harlow, 1950, Harlow et al., 1950). Exploration in chimpanzees has been studied by Welker (1956): in this work several pairs of objects were put in front of the chimpanzees to understand their preferences for one or the other. In this way, size, brightness, and heterogeneity were shown to be important features for eliciting interest towards the stimuli. Chimpanzees also spent a greater time exploring objects that could be moved, changed, or could emit sound and light. Aside the perceptual features of objects, it has also been suggested that interest and interaction can be strengthened by the opportunity to exert control over the environment (Glow et al., 1972; Glow and Winefield, 1978).

In order to study which is the driving force that shapes exploratory behaviors underlying learning processes in humans and non-human primates, we designed a new mechatronic tool for behavioral analysis called “Mechatronic Board” (or

“board” for ease of reference). In the construction of the board we focussed our attention on the experimental needs related to tests involving children, the New World tufted capuchin monkeys (*Cebus Apella*), and humanoid robots. To show an example of the use of the board we illustrate here some pilot experiments involving these two types of participants and directed to test if the free exploration of the board driven by intrinsic motivations allows the participants to acquire actions and action-outcome associations that improve their capacity to solve subsequent biologically-relevant tasks (Baldassarre, 2011). The pilot experiments run with the board and involving children and monkeys are being modelled with biologically-constrained computational models tested in the humanoid robot iCub (Natale et al., 2012) engaged with the same board used to test children. The aim of these models is to understand the computational and biological mechanisms underlying the behaviors and the learning processes observed in the real participants.

2 The Mechatronic Board

The Mechatronic Board is an innovative device specifically designed for interspecies comparative research on intrinsically motivated cumulative learning in children and non-human primates. The board can be also used to test computational models with humanoid robots. This platform has been designed to be modular and easily reconfigurable, allowing to customize the experimental setup according to different protocols devised for children and monkeys (Taffoni et al., 2012). The mechatronic board is the result of a multidisciplinary design process, which has involved bio-engineers, developmental neuroscientists, primatologists and roboticists to identify the main requirements and specifications of the platform. The main requirements, which guided the design and fabrication of the board, are as follows:

- To allow the accomplishment of experiments involving intrinsic and/or extrinsic motivations, that is, respectively, curiosity driven and rewarded actions. Moreover, to allow the learning of actions in a cumulative learning fashion.
- To embed non-intrusive technologies, and to be formed by elements that are ecological and small/light enough to be suitably manipulated by capuchin monkeys and children.
- To be equipped with instrumented interchangeable objects stimulating different kinds of manipulative behaviors so to allow recording several kinds of actions (e.g., rotations, pushing, pulling, repetitive hand movements, button pressing, etc.).
- To record synchronized multimodal information for behavioral analysis and to allow the generation of several different stimuli: visual, acoustic, and tactile.
- To allow the performance of a set of reprogrammable actions, and to allow rewarding them (e.g., food for monkeys; small toys or stickers for children); to allow the participants to see such rewards and at the same time to prevent

their retrieval if necessary (done with automatically closing/opening boxes with transparent covers).

- To be made of materials, mechanisms, and electronic components robust enough to resist actions of monkeys and children.
- To prevent any manipulation or interaction which could be potentially dangerous for the participants or detrimental for the board.

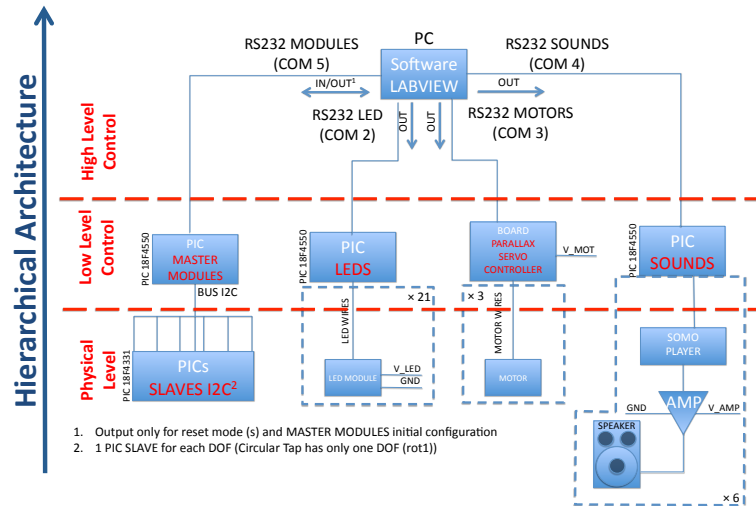


Fig. 1. Hierarchical architecture of the board: physical level made by the interfaces with participants; local low-level control based on microcontrollers; high level control running on a remote laptop.

To easily reconfigure the experimental setup responding to the requirements detailed above, a hierarchical *three-level control architecture* was chosen (see Fig. 1). The *physical level* is made by the interfaces that participants can directly interact with: modules and rewarding mechanisms. This level is mechanically and electronically decoupled by the other higher levels allowing, on the one hand, an easy change of mechatronic modules, and, on the other, an improvement in the robustness of the apparatus. The microcontroller-based *middleware level control* manages low level communication with mechatronic modules, reward mechanisms, and audio-visual stimuli. The *high level control* is a control program

running on a remote laptop which allows supervising the acquisition of data and programming the *arbitrary association between actions and outcomes*.

The mechatronic board has been designed and built in two versions for experiments with capuchin monkeys and children. The two versions of the board are slightly different to take into account the differences between the two groups of participants. The monkey version of the board is heavier, bigger, waterproof (as monkeys could urinate on it), made of non-varnished materials (as monkeys like to remove the paint with their teeth and nails), robust enough to resist to actions such as hitting, rubbing, biting. The children version of the board is similar to the monkey version, but is scaled in dimension and mainly made of wood (see Fig. 2).

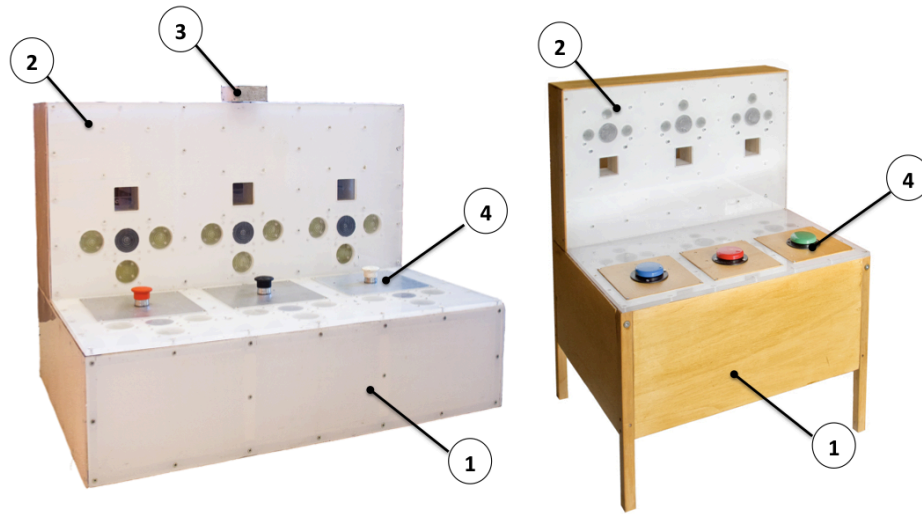


Fig. 2. The mechatronic board for monkeys (left) and children (right) : (1) planar base; (2) reward releasing unit (the squares are boxes that can open to deliver the reward, the nine grey circles are coloured lights, the three dark grey circles are loudspeakers); (3) local wide-angle camera (only in the monkey board); (4) mechatronic modules: in this setting simple pushbuttons. The stimuli/reward system is not visible by the participants, and it controls the aperture and closure of the reward boxes as well as the visual and acoustic stimuli.

Both versions of the board are formed by the following components (Fig. 2):

1. One planar base (800x600x200 mm for the monkeys board, 650x500x450 mm for the children board): it is provided of three slots (200x200 mm; 180x180 mm) where the buttons or the different mechatronic modules can be easily plugged in.

2. The reward releasing unit (800x200x400 mm; 650x120x400 mm): it is mounted on the back area of the planar base and contains the reward boxes where rewards are placed by the experimenter. The boxes are made by transparent material so that the participants can always see inside them. The rear side of the board is provided of suitable openings that allow the experimenter to easily insert the reward in the boxes.
3. A local video-camera: embedded on the top of the reward releasing unit in the monkey version, and external for children. This camera allows recording videos of the work space during the experiments.
4. Pushable buttons and mechatronic modules: each of them is provided with a specific set of sensors and a local microcontroller unit that sends data to the microcontroller-based middleware level through a communication bus (I2C bus). Each module is identified by a hardware address, which guarantees the modularity and the reconfigurability of the system allowing to easily collect data from the different peripherals. For the mechatronic modules, only optical sensors were used in order to physically separate electronics and physical interfaces in order to avoid any direct interaction between participants (monkeys or children) and the electronics of the board. The current architecture allows reconfiguring the platform by substituting the modules (a total of three modules can be plugged at the same time); newly designed modules can also be plugged in as long as they have a unique I2C address.

The board is currently equipped with a set of button modules and three complex mechatronic modules. The button modules allow the detection of a simple pressing action or they can be programmed to respond to more complex interactions such as multiple consecutive presses or a hold press (the time interval can be arbitrarily set by the high level control system). There are also other modules that support the execution of more sophisticated actions. The first mechatronic module, called Circular Tap, measures rotations and vertical translation of about 30 mm. The second one, called Fixed Prism, allows to assess horizontal rotation and translation. The third one, called Three-Degree-of-Freedom Cylinder (3 Dof cylinder), records the movements during the interaction with three different affordances. In the 3 Dof cylinder the effect of interaction can be direct, if the participant rotates the central cylinder or translates it using the horizontal handle, or mediated by an inner mechanism, which translates the rotation of the lateral wheel in an horizontal translation of the cylinder along its main axis. Fig. 3 shows the affordances and the degrees of freedom of the three mechatronic modules.

5. Stimuli and reward system: the whole platform is provided with a set of different stimuli (acoustic and visual) to provide various sensory feedbacks associated to the manipulation of the mechatronic objects. The stimuli come both from the mechatronic objects (object stimuli) and from the reward releasing boxes (box stimuli). The acoustic stimuli are managed by a low-level sound module (Somo-14D manufactured by 4D Systems) that can playback a set of pre-stored audio files; the files that can be used during the experiments can be chosen among a bigger database of natural and artificial sounds. The visual stimuli consist of a set of 21 independent multicoloured

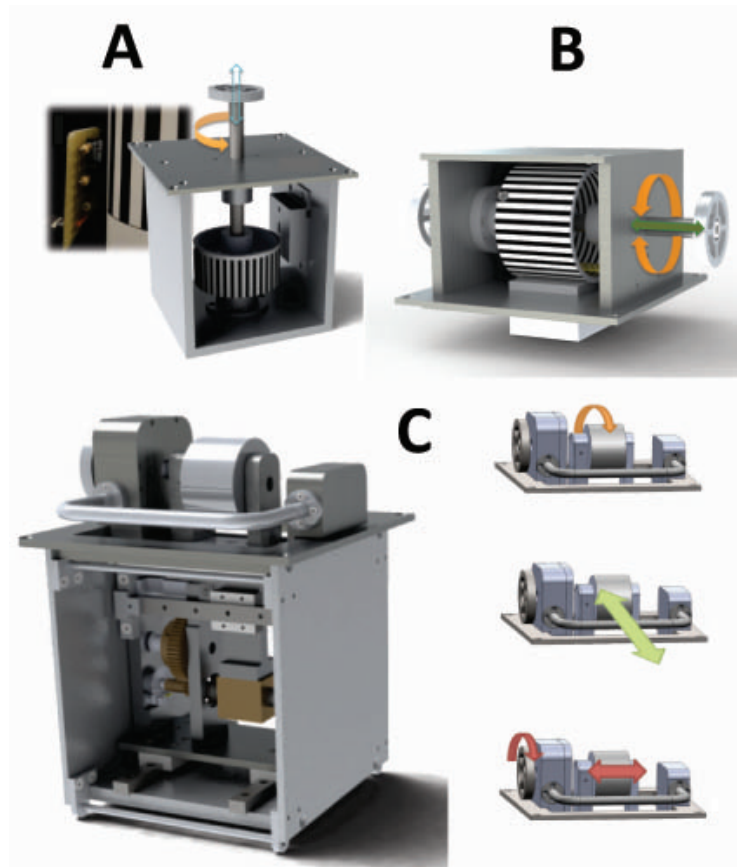


Fig. 3. Mechatronic modules: (A) Circular tap: overall layout and a detail of encoder electronics for rotation measurement; (B) Fixed prism: the frontal wall has been removed allowing to see inner mechanism; (C) 3-Dof cylinder: the overall layout is shown on the left, whereas the degrees of freedom on the right (adapted from [Taffoni et al., 2012](#)).

lights. The actions on the mechatronic objects produce the activation of the audio-visual stimuli and/or the opening of the reward boxes, as defined by the experimental protocol. The reward system is conceived so that the participant can retrieve the reward only if he/she performs the correct action on the mechatronic modules. The reward releasing mechanism was designed to be not backdriveable (so that the participant cannot force the opening; see Fig. 4).

A Parallax Continuous Rotation Servo motor (maximal torque: 0.33 Nm) has been used to drive the box opening mechanism. The motor is coupled to the sliding door by a worm-wheel low efficiency mechanism ($\eta_{tot} = 0.3$). The

low torque of the motor and the low efficiency of the transmission makes the mechanism not harmful if the participants' hand is caught in the sliding door; furthermore, since the mechanism is not backdriveable, it does not allow the participant to force the opening of the sliding door. The action-outcome associations are managed by the high-level control system and are fully programmable according to the experiment requirements.

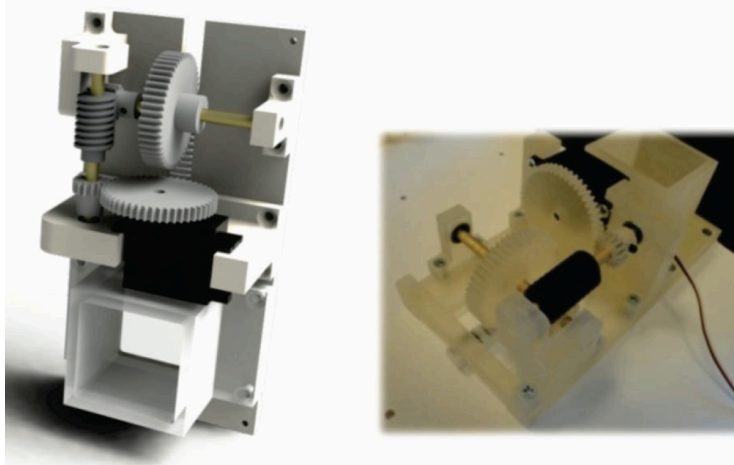


Fig. 4. Reward releasing mechanism. Left: rendering of the mechanism. Right: the developed mechanism.

All the electronics of the microcontroller-based middleware level have been integrated in a single motherboard that could be embedded into the planar base. The motherboard was connected to the audio/video stimuli boards and to the mechatronic modules using 10-way flat cables (see Fig. 5)

3 Experiments with children, monkeys, and the iCub robot

Here, we provide an example of experimental use of the Mechatronic Board equipped with pushbuttons. Pilots experiments were carried out at the day-care centre “La Primavera del Campus” (Universita’ Campus Bio-Medico, Rome, Italy) with children aged between 24 and 51 months, at the “Primate Centre” (Institute of Cognitive Sciences and Technologies, CNR, Rome) with a New World primate species, the tufted capuchin monkey (*Cebus Apella*), and at the Laboratory of Computational Embodied Neuroscience (Institute of Cognitive Sciences and Technologies, CNR) with the iCub humanoid robot. The goal of

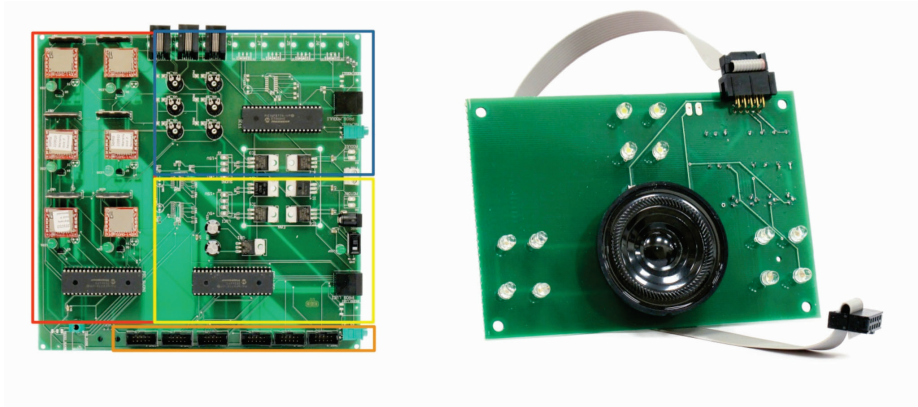


Fig. 5. Left: motherboard with the electronics of the microcontroller-based middleware system: PIC master for mechatronic modules (within the top-right blue rectangle); PIC for Led control (centre right yellow rectangle); PIC for sounds control (left green rectangle); connectors for audio/video stimuli board (bottom orange rectangle). Right: audio-visual stimuli board.

the experiments was to verify if intrinsic motivations can drive monkeys, children and a computational model tested on a humanoid robot to spontaneously explore the board and so acquire knowledge on the possible action-outcome associations that characterize it (e.g., the pressure of certain buttons opens certain boxes). The acquisition of this knowledge is tested in a second experimental phase where the achievement of a certain outcome, for example the opening of a box, is made desirable by associating with it a reward (for example, a food or sticker is inserted into a particular box that can be opened by pressing a certain button). The ease with which a certain outcome is accomplished is a measure of the knowledge acquired on the basis of intrinsic motivations in the previous exploration phase.

3.1 Experimental Protocols

3.1.1 Children The experiments are performed by placing the board in an empty room where the child is introduced by his/her teacher. The teacher invites the child to explore the board by saying “*Look at this new toy. What is this? What can it do?*”, without saying anything about what the board actually does. As mentioned, the experimental protocol is divided in two phases, a training phase and a test phase, and involves two groups of participants, an experimental group and a control group.

During the training phase, which lasts 10 minutes, the children of the experimental group can discover by spontaneous exploration that they can open each box (empty in this phase) by pressing a certain button for more than one

second. When a box is opened, a light inside the box turns on, and a sound of an animal cry is produced (a different one for each button: a rooster, a frog, and a cat). For both groups, the simple press of a button makes the lights close to the button to turn on and produces a single xylophone note (three different notes are set for the three buttons). For both groups, the blue button (Bb) opens the left box (LB), the red button (Rb) the right box (RB) and the green button (Gb) the central box (CB) (see Fig. 6).

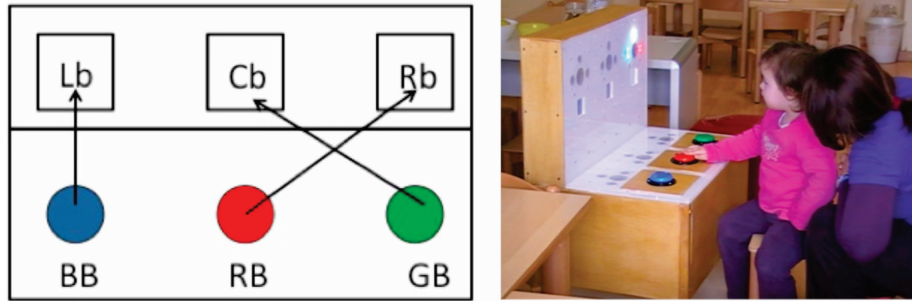


Fig. 6. *Left:* Schematic representation of the arrangement of buttons and their association with boxes from the perspective of the user. *Right:* A snapshot of the experiment during the training phase.

In the test phase, equal for the two groups, the reward (a sticker) is pseudo-randomly placed in one of the three closed boxes where it is clearly visible to the participants. The child is invited to retrieve the sticker without receiving any indication on which action to perform to open the box. As in the training phase the box can be opened by pushing and holding the particular button associated to the box for more than one second. Pressing a button also causes the other stimuli as in the training phase.

Once the participant opens the box and reaches the sticker, this is given to the child as a reward. If the participant does not retrieve the sticker after two minutes, the sticker is moved to the next box. The test phase ends after nine successful openings of the boxes (three for each box) or after 18 minutes.

3.1.2 Monkeys In the experiment with monkeys, the board has three buttons of different colors (white, black, and red) placed about 25 cm apart from one another along the same line (see Fig. 7). The buttons can be discriminated by trichromatic and dichromatic subjects (capuchin monkeys male are all dichromats, whereas females can be either dichromats or trichromats, [Jacobs, 1998](#)). The pressure of each button produces a specific combination of audio and visual stimuli along with the opening of one of the three boxes. Each subject is

separated from the social group it belongs to just before the experimental session. Subjects are not food deprived and water is freely available at all times. The pilot experiment includes two phases as for the children protocol. During the training phase the correct action performed by the subject (i.e., pressing a button at least once) produces a specific combination of audio and visual effects together with the opening of one box. The box does not contain any reward. The training phase lasts 20 minutes.

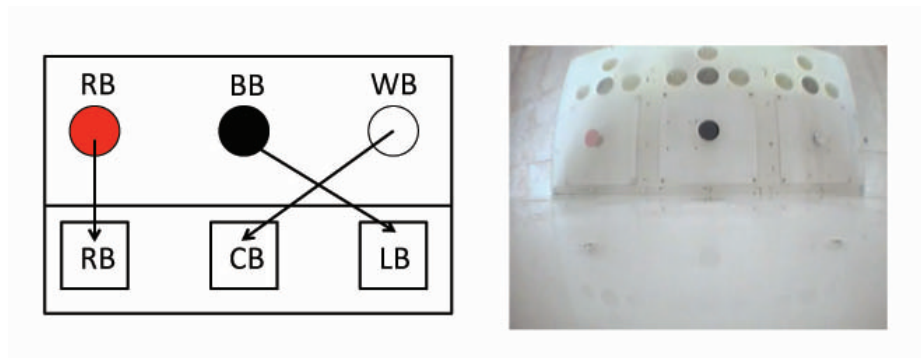


Fig. 7. Configuration of buttons and their association with boxes from the monkey's perspective (adapted from [Taffoni et al., 2012](#)).

In the test phase, the reward (one peanut kernel) is located in one of the three boxes in clear view of the subject. The reward can be obtained by pressing the associated button. Each subject receives nine trials. The reward position was balanced across boxes. The test phase ends after nine trials or when 40 minutes elapse. For all subjects, the white button (WB) opens the central box (CB), the black button (BB) the left box (LB) and the red button (RB) the right box (RB) (see Fig. 7). Thus, the spatial relation between button and associated box is crossed for WB and BB and frontal for RB. The pilot experiment is videotaped by a camera (Sony Handycam, DCR-SR35) and by the camera embedded in the board. The ELAN software was used to synchronize the videos obtained by the two cameras.

3.1.3 Robot The iCub robot was used to test a computational model of the board experiment run with monkeys and children (see Fig. 8). The model was built in order to formulate an operational hypothesis on the neural and computational mechanisms that might underlie the behaviors observed in the experiments with real participants.

The test with the robot is divided in two phases as the experiments involving the real participants. In the learning phase, the robot can press any button of

the board. The pressure of a button causes the opening of a box. For simplicity, the buttons and boxes spatially correspond, and no sound nor light is caused by the button pressure. The opening of a box causes a surprising, unexpected event that leads the robot to learn the action-outcome association between the action executed and its effect (box opening). In the test phase, one particular outcome (e.g., box 1 opening) is given a high value (to this purpose, for now the neural representation of the outcome is manually activated). As a consequence, if the robot has suitably learned the action-outcome associations related to that outcome during the learning phase, it is expected to be able to immediately recall the correct action to obtain the reward.



Fig. 8. The iCub robot that interacts with the board.

We now explain the model architecture (see Fig. 9) and functioning (see [Baldassarre et al., submitted](#) for further details on the model and its results). The empirical experiments show that when monkeys and children face the board experiment for the first time, they are already endowed with a suitable action repertoire that allows them to explore the board by performing quite sophisticated actions: for example, they focus their attention on various parts of the board quite efficiently and execute various actions such as “reach”, “touch”, “press”, “scratch”, etc. For this reason, the computational model is formed by two main components: a sensorimotor component (SMc), which learns and executes the movements needed to implement arm and eye actions, and a decision making component (DMc), which learns to select and trigger the execution of such actions in the correct context.

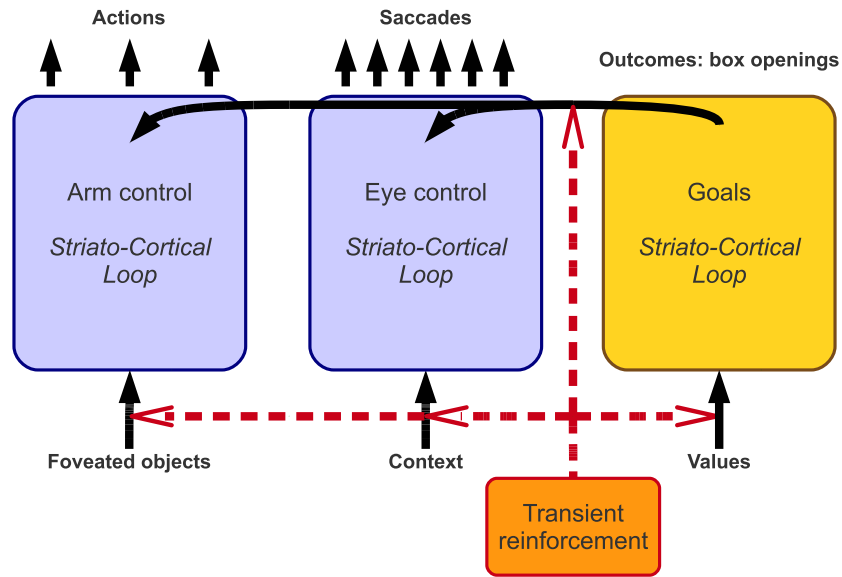


Fig. 9. The architecture of the computational model controlling the robot. Blue boxes: striato-cortical loops involving premotor and parietal cortex and controlling the arm and eye movements. Yellow box: loop selecting desired outcomes (goals). Red box: dopamine centres.

Sensorimotor component. As mentioned above, this component is responsible for executing the movements that implement the actions (such as reach and press) when they are selected by the DMc. This component might correspond to cortico-cortical sensorimotor pathways of dorsal neural pathways of the brain (Caligiore et al., 2010; Goodale and Milner, 1992). The actions are acquired with suitable learning processes before the system tackles the two phases of the experiment, so mimicking the acquisition of skills by monkeys and children before the experiment with the board. The actions used by the model were six actions for the eye (“look at button x”, and “look at box y”, where x and y were respectively the three buttons and the three boxes) and three arm actions (“press the looked object”, and other two dummy actions introduced to test the learning capabilities of the system). Each action was learned on the basis of a sensorimotor mapping where the sensory and motor spaces were represented with maps of receptive or motor fields (cf. Lee et al., 2007). Given the focus of this paper on the high-level cognitive aspects of the experiments with children and monkeys, the SMc of the system is no further discussed here.

Decision making component. This component is responsible for deciding the actions to perform based on mechanisms putatively implemented by three *striato-cortical loops* of brain (Yin and Knowlton, 2006). Striato-cortical loops are formed by basal ganglia, sub-cortical nuclei responsible for learning to select actions in the correct context, and various frontal cortical regions, responsible for performing different actions such as reach, press, look at a particular place in the environment, and for encoding action outcomes. In the model, two loops involving the premotor and parietal cortex are responsible for selecting respectively eye actions (e.g., “look at a particular button or box”) and arm actions (e.g., “press” the object you are looking). The third loop, involving the prefrontal cortex, is responsible for encoding action outcomes (e.g., “box 1 is opening”; cf. Miller and Cohen, 2001).

During the learning phase, the model initially selects and performs random actions by looking various parts of the board and by executing arm action on them. Based on this exploration, the model learns to associate particular experienced outcomes with particular actions, for example the fact that when button 1 is pressed box 1 opens. This learning process involves the formation of connections between the loop encoding outcomes and the other two loops selecting actions based on a Hebbian rule. This learning process is driven by learning signals putatively corresponding to the production of dopamine, a neuromodulator playing an important role in trial-and-error learning processes taking place in striato-cortical loops and in frontal cortex (Houk et al., 1995). Following Redgrave and Gurney (2006), in the model these learning signals are produced by the perception of the sudden opening of boxes: this surprising event activates another component of the system, putatively corresponding to the superior colliculus, that in turn causes the dopamine signal. The dopamine signal also drives a second learning process involving the striato-cortical loops and this causes the model to repeat the last performed actions several times. This facilitates the

learning process that forms the action-outcome association mentioned above (this mechanism is called “repetition bias”; Redgrave et al., 2012).

An important aspect of the model is the fact that the dopamine learning signal is *transient*: it progressively fades away if the surprising event (e.g., opening of box 1) is experienced several times (Mirolli et al., submitted). In turn, the decrease of the leaning signal causes an unlearning process within the striato-cortical loops and this cause a decrease of the tendency to produce the actions that lead to the outcome. In this way the system can focus its activity on other action-outcomes to be learned.

3.2 Preliminary results

3.2.1 Children Six children aged between 24 and 51 months were involved in the experiment. All children were identified as right-handed by their teachers (in future experiments, we plan to assess manual preference using the Oldfield inventory). Children were age-matched according to three age groups and assigned to the experimental group (EXP) or the control group (CTRL) (see table 1).

- **Age group 1**, two children, mean age 24 months
- **Age group 2**, two children, mean age 31.5 months
- **Age group 3**, two children, mean age 49.5 months

Table 1. Children involved in the pilot experiment.

Subject	Group	Age [Months]
CBM08	EXP	24
CBM06	CTRL	24
CBM11	EXP	33
CBM09	CTRL	32
CBM17	EXP	48
CBM19	CTRL	51

During the training phase both CTRL and EXP groups were exposed to the board for 10 minutes. As illustrated in section 3), the box openings were disabled for the CTRL group. Table 2 summarises the interaction of each participant with the board during this phase.

Younger participants (age group I and II) seem to prefer middle and right presses (respectively Rb and Gb) possibly because participants are right-handed. Such preference is not observed in children of group age III. Interestingly, there is a progressive increase of the number of pushes of the Lb from the younger to older age. No significant differences in terms of number of correct actions were observed between the experimental and control groups.

Table 2. Number of interactions of each participant with the board during the training phase. “*” refers to the percentage of button pushes lasting longer than 2s.

Participant	Presses	Correct Actions*	Activation box 1(%)	Activation box 2(%)	Activation box 3(%)
CBM08	142	57	14.03	42.11	44.86
CBM06	292	27	18.51	33.33	48.16
CBM11	92	19	21.05	36.84	42.11
CBM09	102	59	25.42	30.51	44.07
CBM17	239	49	12.24	48.98	38.78
CBM19	365	36	36.11	30.56	33.33

To assess the transfer of motor skills into the new context, during the test phase participants were asked to retrieve a sticker inserted into one of the three boxes. Nine stickers were used (three for each box), and the insertion order was random (see section 3). Participants in the experimental group were found to retrieve a higher number of rewards. The training effect increases dramatically with age (Fig. 10).

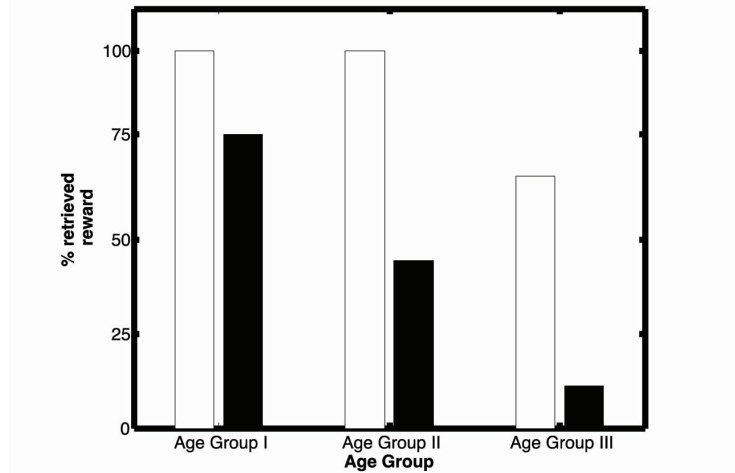


Fig. 10. Retrieved rewards (%): comparison between EXP (white) and CTRL (black) groups.

To assess if participants have learned the spatial relationship between buttons and boxes we defined a *spatial relationship index* (SRI) as:

$$SRI = \frac{\text{Number of correct pushes}}{\text{number of total pushes per trial}} \quad (1)$$

According to this index, if a participant presses only the button which controls the opening of the box where the reward is placed, such index will tend to one; if a participant presses randomly all the buttons such index will tend to 0.33; if a participant learns a wrong relationship such index will tend to zero.

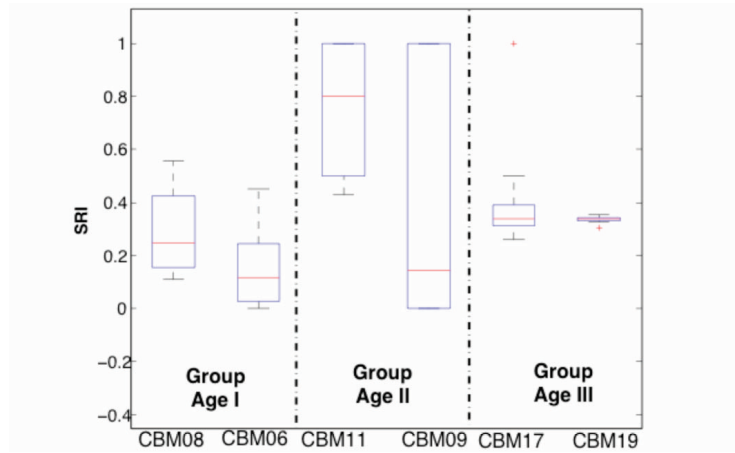


Fig. 11. Spatial relationship index measured for each participant.

Table 3. Differences in the SRI between EXP and CTRL group

Subject	SRI SIMPLE	SRI CROSSED
	[Mean \pm SD]	[Mean \pm SD]
CBM08	0.39 \pm 0.24	0.24 \pm 0.10
CBM06	0.20 \pm 0.22	0.14 \pm 0.17
CBM11	0.95 \pm 0.1	0.62 \pm 0.23*
CBM09	1	0.08 \pm 0.12*
CBM17	0.34 \pm 0.01	0.45 \pm 0.28
CBM19	0.32 \pm 0.01	0.33 \pm 0.01

Fig. 11 presents the boxplot of the SRI values for the six participants involved in the experiment. Red lines represent the median value of the index. Younger

participants (Age Group I) do not seem to have learnt the spatial relationship between buttons and boxes; in Age Group II, the two medians seem to suggest that there could be a difference between the CTRL and the EXP group: in particular the CTRL group seems to behave in a random way. Children of Age Group III do not show any difference and seem to act in a random way. This suggests that the children in the EXP group have learned that pressing a button for a longer period opens the boxes and are able to exploit this knowledge to retrieve the reward. However, they did not discover the spatial relationship between the buttons and the boxes. Considering separately the trials where reward was placed in boxes CB and RB cross related to buttons opening them, children of Age Group II seem to show a statistically significant difference ($*F(1,10)=25.72, p=0.0007$) in SRI between CTRL and EXP.

3.2.2 Monkeys The subjects of the pilot experiments were 3 adult capuchin monkeys (Pedro, Robiola and Robin Hood). During the training phase, two subjects contacted the board within a few sec (Robiola, 6 sec and Robin Hood, 37 sec) whereas Pedro took much longer (6 min and 27 sec). Robiola performed her first push directed toward a button 1 min and 15 sec after the beginning of the trial, whereas the other subjects never did it. Robiola pressed all the buttons at least twice, for a total of 14 pushes. The average time during which she held the button pressed was 0.17 sec (SE: ± 0.008). The average time of contact of the subjects with the board was 5 min and 5 sec (Robiola: 10 min and 38 sec; Pedro: 3 min and 55 sec; Robin Hood: 3 min and 11 sec). Each button was manipulated for a mean of 15.55 sec (SE: ± 2.02) during Phase 1. The overall mean scratching rate, (used as a behavioral measure of stress) occurred at 0.4 events/min (SE: ± 0.02).

During the test phase seeing a reward in one of the boxes prompted subjects attention towards it and increased his/her motivation to manipulate the board. Capuchins readily visually explored the baited box; this behavior was much more frequent than in the previous phase (Pedro: 170 times, Robin Hood: 132, Robiola: 20). Indeed, subjects spent much more time on the board (Mean \pm SE: 19 min and 10 sec ± 2.76) and manipulated each button much longer (Mean \pm SE: 40 sec ± 8.03). Scratching occurred at a higher rate than in Phase 1 (Mean \pm SE: 0.6 events/min ± 0.05).

Table 4 shows for the three box-button associations the mean number of incorrect responses before pushing the correct button, the number of times each button is pressed and the mean holding time of each button. Overall, the frontal association (right box-red button) had a mean number of errors similar to the left box-black button crossed association, whereas the other crossed association (central box-white button) scored a higher level of errors (see also Fig. 2). The black button located in the central position (operating the left box) was pressed almost twice the other two buttons, therefore increasing the probability to open the left-box. Consequently, the comparison between frontal and crossed associations should be carried out by comparing the performances in the right and in the central box. Since the mean number of errors per trial per subject was 1.2 (right

box) and 3.7 (central box), we suggest that spatial proximity plays a primary role in learning an association between action and outcome. The SRI for the three subjects involved in the experiment highlights an easier understanding of direct spatial relation ($SRI = 0.76$) with respect to the crossed ones ($SRI = 0.51$ for the black button-left box and $SRI = 0.34$ for white button-central box)

Table 4. association between boxes and buttons

	Left Box Black Button	Central Box White Button	Right Box Red Button
Mean number of pushes per subject per trial \pm SE	1.9 ± 0.8	0.8 ± 0.3	1 ± 0.25
Mean number of incorrect responses per subject per trial \pm SE	1.2 ± 0.2	3.7 ± 0.7	1.2 ± 0.3
Mean holding time per subject per trial \pm SE	0.2 ± 0.05	0.25 ± 0.03	0.3 0.11

3.2.3 Robot Fig. 12 shows the behavior of the computational model tested with the iCub robot during the training phase and during the test phase. Each histogram bar shows the percent of times the system performs the three action compounds involving an action of the eye and one of the arm: “look at button x & press button x” indicated with different colours for the three buttons. The graph also reports the other six action compounds of the type “look at button x & arm performs dummy action y on button x” (recall that there are two dummy actions). Notice how during training the system initially performs actions randomly, then focuses on performing the correct arm action (press) on button 2, then focuses on performing the correct action on button 1, then focuses on performing the correct action on button 3, and finally performs again random actions as at the beginning. This result shows that the repetition bias allows the model to focus its attention on a particular button for a certain period of time, and then to repeat the press action several times. This leads the model to quickly learn to associate the outcome (box opening), represented in the prefrontal cortex loop, to the eye action of looking at the pressed button, represented in the eye loop, and to the arm action of pressing that button, represented in the arm loop.

The last three histogram bars of Fig. 12 indicate the behavior of the model when, during the testing phase, each of the three goals is manually activated in prefrontal cortex for a certain period of time (the three bars refer to the three different goals). The graph shows that when an outcome is activated the system is capable of recalling the suitable action compound that allows the system to achieve that outcome. For example, when the outcome “opening of box 1” is activated, the system looks at button 1 and presses it.

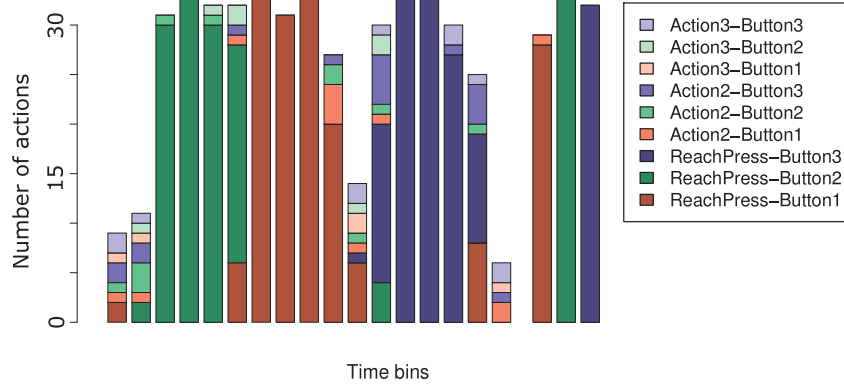


Fig. 12. Behavior of the model during the training phase and the test phase. The y-axis shows the actions performed during time bins (x-axis) of 5% of the total time of the experiment (120 min). The y-axis represents the number of selections and executions of the three arm actions on the three buttons when these are looked by the eye (nine combinations in total). The actions directed to the boxes are omitted for simplicity. The first 17 bins refer to the learning phase. The last three bins refer to the test phase: for each bin one of the three goals was manually activated. Notice how the model initially performs actions randomly, then focuses on pressing one button (from bin 2 to bin 6), then on pressing a second button (from bin 7 to bin 10), then on pressing the last button (from bin 12 to bin 15), and finally on exploring again (last bin). Also notice how the system executes the correct actions during the last three bins of the test, thus demonstrating goal-directed behavior.

4 Discussion

The tendency of young children to explore more frequently central and right pushbuttons could affect the way in which they learn new skills. Preliminary results seem to suggest that children who were given the chance of discovering a new skill based on intrinsic motivations had a higher chance of employing that skill to later obtain valuable outcomes. However, neither the experimental nor the control groups did learn more complex spatial relationships. These results suggest to focus on an age range between 36 and 48 months in order to avoid problems related to motor control development.

Regarding monkeys, the results suggest that capuchin monkeys were not very interested in the buttons during the learning phase, whereas their interest toward the board significantly increased during the test phase due to the view of the reward. In this phase the board triggered a variety of behaviors, such as visual exploration, prolonged contact with the apparatus, and pressing the buttons.

These behaviors may eventually lead capuchins to learn specific action-outcome associations. The association between boxes and buttons in the crossed condition was perceived by monkeys as more challenging than the frontal association, while there was a strong bias toward the central black button that decreased the number of errors when opening its associated box (the crossed left box). Although we did not collect specific data on subjects' positions on the board, this effect was probably due to the fact that monkeys spent more time at the centre of the board, where the black button was placed, than at the left and right sides. Overall, the results highlight the role of extrinsic rewards and spatial proximity as critical factors affecting capuchins' learning processes and point out the importance of choosing suitable objects that promote interest and manipulation. Very likely, buttons were too simple and afford only the action of pressing. We may thus suggest that the use of the mechatronic board equipped with modules rather than buttons would likely elicit an increase in capuchins' interest toward the apparatus.

The results obtained with the computational model and the robot have shown that the model is able to focus on performing actions that produce interesting events on the board. This focusing is transient and when the system has acquired experience on one aspect of the board it gets bored and starts to explore other aspects. Moreover, the model is capable of recalling the execution of suitable actions when a goal representation is activated by some internal event based on the experience acquired with intrinsic motivations. For now the mechanism that activates the goals is hardwired but in the future it will be substituted by extrinsic motivation mechanisms based on the amygdala (a set of brain nuclei that play a key role in motivations and the assignment of subjective value to stimuli). The added value of the model resides in the fact that it explains the target phenomena by furnishing detailed computational mechanisms that might underlie the target behaviors. Moreover, the overall architecture of the model and the functions ascribed to its components has been constrained on the basis of neuroscientific knowledge on the relevant parts of the brain that are known to play important functions for the production of such behaviors. In this respect the model, although preliminary in many aspects, represents an important base for the future development of further models and a sound understanding of how intrinsic motivations drive learning and exploitation of actions in real organisms. Aside its scientific value, this knowledge can also contribute to guide the design of fully autonomous learning robots.

5 Conclusions

In this work we presented a new mechatronic platform, called *Mechatronic Board*, for studying intrinsically motivated learning in children, monkeys, and robots. The paper has presented a thorough discussion of the main design principles and implementation features of the platform, and has illustrated three examples of how it can be exploited to investigate intrinsic motivations based on three pilot experiments run with children, monkeys, and the humanoid robot iCub.

The preliminary data so obtained clearly indicate that the platform can be effectively used for the behavioral analysis of humans, non-human primates and robots engaged in acquiring skills based on intrinsic motivations. Despite these preliminary experiments were carried out using the platform equipped with only with buttons, more challenging mechatronic objects allowing different possibility of interaction have been designed and will be used in future comparative studies. An important issue for future research is to develop an experimental paradigm involving the board with which to investigate the different kinds of intrinsic motivations that different subjects (e.g. monkeys and children) have. In particular, we are interested in developing a paradigm with which to experimentally assess the presence, in animals and humans, of the competence-based intrinsic motivations (Miroli and Baldassarre, 2012) that have been suggested by recent computational models (Baldassarre and Miroli, 2012; Schembri et al., 2007a,b,c; Stout and Barto, 2010).

6 Acknowledgements

This research was supported by the European Community 7th Framework Programme (FP7/2007-2013), “Challenge 2 - Cognitive Systems, Interaction, Robotics”, Grant Agreement No. ICT-IP-231722, Project “IM-CLeVeR - Intrinsically Motivated Cumulative Learning Versatile Robots”. It was also supported by the Italian Ministry of University and Research, FIRB Research Program 2006, no. RBAP06SPK5.

Bibliography

- Baldassarre, G. (2011). What are intrinsic motivations? a biological perspective. In Cangelosi, A., Triesch, J., Fasel, I., Rohlfing, K., Nori, F., Oudeyer, P.-Y., Schlesinger, M., and Nagai, Y., editors, *Proceedings of the International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2011)*, pages E1–8. IEEE, Piscataway, NJ.
- Baldassarre, G., Mannella, F., Fiore, V., Redgrave, P., Gurney, K., and Mirolli, M. (Submitted). Intrinsically motivated action-outcome learning and goal-based action recall: A system-level bio-constrained computational model. *Neural Networks*.
- Baldassarre, G. and Mirolli, M. (2012). Deciding which skill to learn when: Temporal-difference competence-based intrinsic motivation (td-cb-im). In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).
- Caligiore, D., Borghi, A., Parisi, D., and Baldassarre, G. (2010). Tropicals: A computational embodied neuroscience model of compatibility effects. *Psychological Review*, 117(4):1188–1228.
- Glow, P., Roberts, J. S., and Russell, A. (1972). Sound and light preference behaviour in naïve adult rats. *Australian Journal of Psychology*, 24:173–178.
- Glow, P. and Winefield, A. H. (1978). Response-contingent sensory change in a causally structured environment. *Animal Learning and Behaviour*, 6:1–18.
- Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci*, 15(1):20–25.
- Harlow, H. F. (1950). Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *Journal of Comparative and Physiological Psychology*, pages 289–294.
- Harlow, H. F., Harlow, M. K., and Meyer, D. R. (1950). Learning motivated by a manipulation drive. *Journal of Experimental Psychology*, 40:228–234.
- Houk, J. C., Adams, J. L., and Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In Houk, J. C., Davids, J. L., and Beiser, D. G., editors, *Models of Information Processing in the Basal Ganglia*, pages 249–270. The MIT Press, Cambridge, MA.
- Jacobs, G. (1998). A perspective on colour vision in platyrrhine monkeys. *Vision Research*, 38:3307–3313.
- Kaplan, F. and Oudeyer, P. (2007). In search of the neural circuits of intrinsic motivation. *Frontiers in neuroscience*, 1:225–36.

- Klemke, E. D., editor (1980). *Introductory Readings in the Philosophy of Science*. Prometheus Books, New York.
- Lee, M. H., Meng, Q., and Chao, F. (2007). Staged competence learning in developmental robotics. *Adaptive Behavior*, 15(3):241–255.
- Lockmann, J. J. (2000). A perception-action perspective on tool use development. *Child Development*, 71(1):137–144.
- Martin, P. and Bateson, P., editors (1998). *Measuring Behaviour: an introductory guide*. Cambridge University Press, Cambridge.
- Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*, 24:167–202.
- Mirolli, M. and Baldassarre, G. (2012). Functions and mechanisms of intrinsic motivations: The knowledge versus competence distinction. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).
- Mirolli, M., Santucci, V. G., and Baldassarre, G. (Submitted). Phasic dopamine as a prediction error of intrinsic and extrinsic reinforcements driving both action acquisition and reward maximization: A simulated robotic study. *Neural Networks*.
- Natale, L., Nori, F., Metta, G., Fumagalli, M., Ivaldi, S., Pattacini, U., Randazzo, M., Schmitz, A., and Sandini, G. (2012). The icub platform: a tool for studying intrinsically motivated learning. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag, Berlin. (this volume).
- Redgrave, P. and Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nat Rev Neurosci*, 7(12):967–975.
- Redgrave, P., Gurney, K., Stafford, T., Thirkettle, M., and Lewis, J. (2012). The role of the basal ganglia in discovering novel actions. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag, Berlin. (this volume).
- Schembri, M., Mirolli, M., and Baldassarre, G. (2007a). Evolution and learning in an intrinsically motivated reinforcement learning robot. In Almeida e Costa Fernando, Rocha, L. M., Costa, E., Harvey, I., and Coutinho, A., editors, *Advances in Artificial Life. Proceedings of the 9th European Conference on Artificial Life (ECAL2007)*, volume 4648 of *Lecture Notes in Artificial Intelligence*, pages 294–333. Springer Verlag, Berlin. Lisbon, Portugal, September 2007.
- Schembri, M., Mirolli, M., and Baldassarre, G. (2007b). Evolving childhood’s length and learning parameters in an intrinsically motivated reinforcement learning robot. In Berthouze, L., Dhristiopher, P. G., Littman, M., Kozima, H., and Balkenius, C., editors, *Proceedings of the Seventh International Con-*

- ference on Epigenetic Robotics*, volume 134, pages 141–148. Lund University, Lund.
- Schembri, M., Mirolli, M., and Baldassarre, G. (2007c). Evolving internal reinforcers for an intrinsically motivated reinforcement-learning robot. In Demiris, Y., Mareschal, D., Scassellati, B., and Weng, J., editors, *Proceedings of the 6th International Conference on Development and Learning*, pages E1–6, London. Imperial College.
- Stout, A. and Barto, A. G. (2010). Competence progress intrinsic motivation. In Kuipers, B., Shultz, T., Stoytchev, A., and Yu, C., editors, *IEEE International Conference on Development and Learning (ICDL2010)*. IEEE, Piscataway, NJ.
- Taffoni, F., Vespignani, M., Formica, D., Cavallo, G., Polizzi di Sorrentino, E., Sabbatini, G., Truppa, V., Mirolli, M., Baldassarre, G., Visalberghi, E., Keller, F., and Guglielmelli, E. (2012). A mechatronic platform for behavioral analysis on nonhuman primates. *Journal of Integrative Neuroscience*, 11(1):87–101.
- Thelen, E. and Smith, L. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Boston, MA, USA, MIT Press.
- Wasserman, E. and Zentall, T. (2006). *Comparative Cognition: Experimental Explorations of Animal Intelligence*. Oxford University Press, New York.
- Welker, W. L. (1956). Some determinants of play and exploration in chimpanzees. *Journal of Comparative Physiological Psychology*, 49:84–89.
- White, R. W. (1959). Motivation reconsidered: the concept of competence. *Psychological Review*, 66:297–333.
- Yin, H. H. and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nat Rev Neurosci*, 7(6):464–476.