Experimental Brain Research manuscript No.
(will be inserted by the editor)

Development of goal-directed action selection guided by intrinsic motivations: an experiment with children

Fabrizio Taffoni* · **Eleonora Tamilia*** · **Valentina Focaroli** · **Domenico Formica** · **Luca Ricci** · **Giovanni Di Pino** · **Gianluca Baldassarre** · **Marco Mirolli** · **Eugenio Guglielmelli** · **Flavio Keller**

Received: date / Accepted: date

Abstract Action selection is extremely important, particularly when the accomplishment of competitive tasks may require access to limited motor resources. The spontaneous exploration of the world plays a fundamental role in the development of this capacity, providing subjects with an increasingly diverse set of opportunities to acquire, practice and refine the understanding of action-outcome connection. The computational modeling literature proposed a number of specific mechanisms for autonomous agents to discover and target interesting outcomes: intrinsic motivations hold a central importance among those mechanisms. Unfortunately, the study of the acquisition of action-outcome relation was mostly carried out with experiments involving extrinsic tasks, either based on rewards or on predefined task goals. This work presents a new experimental paradigm to study the effect of intrinsic motivation on action-outcome relation learning and action selection during free exploration of the world. Three- and four-year-old children were observed during the free exploration of a new toy: half of them were allowed to develop the knowledge concerning its functioning; the other half were not allowed to learn anything. The knowledge acquired during the free exploration of the toy was subsequently assessed and compared.

Keywords Intrinsic Motivation · Action Selction · Curiosity · Action-Outcome Contingency · Novelty Detection

*F. Taffoni and E. Tamilia, equally contributed to this work

F.Taffoni (*corresponding*) · E. Tamilia · D. Formica · L. Ricci · E. Gugliemelli
Lab. of Biomedical Robotics and Biomicrosystems,
Università Campus Bio-Medico di Roma,
via A. del Portillo 21, 00128 Rome, Italy
Tel.: +39-06-225419610
E-mail: f.taffoni@unicampus.it

V. Focaroli · F. Keller
Lab. of Developmental Neuroscience and Neural Plasticity
Università Campus Bio-Medico di Roma,
via A. del Portillo 21, 00128 Rome, Italy

G. Di Pino
Institute of Neurology, Fondazione Alberto Sordi - Research Institute for Ageing,
Lab. of Biomedical Robotics and Biomicrosystems,
Università Campus Bio-Medico di Roma,
via A. del Portillo 21, 00128 Rome, Italy

G. Baldassarre · M. Mirolli
Laboratory of Computational Embodied Neuroscience,
Institute of Cognitive Sciences and Technologies, CNR
via S. M. della Battaglia 44, 00185 Rome, Italy

1 Introduction

The fast acquisition of the capacity to interact with the world, solve problems and pursue own personal goals is one of the most astonishing manifestations of human intelligence (Piaget and Cook, 1952, von Hofsten, 2004, Keen, 2011). The mechanisms underlying this process are only partially known, see Gottlieb et al (2013) for a review. The observation of infants makes it quite clear that the ability to perform goal-directed actions develops with a continuous open-ended process. This process leads them to understand how actions can accomplish different goals (von Hofsten, 2004, Smith and Gasser, 2005). In particular, thanks to a free exploration of the world, infants discover the potential changes (or *outcomes*) that their actions can cause in the environment, and register the dependencies between such changes and the performance of specific actions, i.e. *action- outcome contingencies*, (Kenward et al, 2009). Indeed, learning of action-outcome contingencies allows children to

Published as: Taffoni Fabrizio, Tamilia Eleonora, Focaroli Valentina, Formica Domenico, Ricci Luca Di Pino Giovanni, Baldassarre Gianluca, Mirolli Marco, Guglielmelli Eugenio, Keller Flavio (2014) Development of goal-directed action selection guided by intrinsic motivations: an experiment with children Experimental Brain Research, 232(7): 2167-2177
doi: 10.1007/s00221-014-3907-z

select the most appropriate motor program to reach a desired outcome, i.e. a goal (Kenward et al, 2009).

The computational modeling literature has proposed a number of specific mechanisms for autonomous agents to discover and target interesting outcomes (see Botvinick et al (2009), for a brief review). Among these, intrinsic motivations (Baldassarre and Mirolli, 2013) hold a central importance in the independent identification of potentially useful outcomes and self generation of goals. According to Ryan and Deci (2000), intrinsic motivations are defined as the motivations driving an activity for its inherent satisfaction rather than because it is instrumental for the attainment of outcomes having a direct biological value (e.g., the achievement of food or the avoidance of pain). In this view, intrinsic motivations drive children to learn for the sake of experience itself, rather than because of any reward given by an adult or the environment. This perspective is very close to the constructivist approach suggested by Piaget and Cook (1952). As in constructivism, learning has a central role, but while constructivism is mainly concerned with the learning process, intrinsic motivations are related to the particular forces that drive children to learn. Intrinsic motivations may be divided into two main classes: competence-based and knowledge-based intrinsic motivations (Kaplan and Oudeyer, 2007, Mirolli and Baldassarre, 2013). Competence-based intrinsic motivations are linked to an agent's behavior, and in particular to the ability (competence) of the agent to modify the world in certain ways (White, 1959). In the computational models of competence-based intrinsic motivations, the agent is typically rewarded when its ability to accomplish a goal improves, independently from the origin of the goal (Chentanez et al, 2004, Schembri et al, 2007, Baranes and Oudeyer, 2013, Santucci et al, 2013). In contrast, knowledge-based intrinsic motivations are linked to the stimuli the agent perceives and to their relation with the agent's previous knowledge (Berlyne, 1960). Recent studies have suggested to divide knowledge-based intrinsic motivations into two sub-classes: novelty-based and prediction-based (Baldassarre and Mirolli, 2013, Barto et al, 2013). Prediction-based intrinsic motivations come forward when the agent's expectations are not met; they are typically modeled through prediction errors or improvements in prediction errors of an agent's model of the world (e.g. Schmidhuber (1991), Kaplan and Oudeyer (2007), Mirolli and Baldassarre (2013)). Novelty-based intrinsic motivations are elicited by objects, or object combinations, that have not been experienced before and hence are not in the agent's memory. Computational models of this sort are typically based on the detection of anomalous/unfamiliar items (see Nehmzow et al (2013) for a review).

Recent neuroscientific research has started to uncover the neural bases of intrinsic motivation mechanisms. Redgrave and Gurney (2006) have argued that sensory prediction errors related to surprising events cause bursts of dopamine, which lead the basal-ganglia to repeat and refine the action that produced the interesting event (Redgrave et al, 2011). Unfamiliar items or a novel combination of sequences of them seems to be detected by the hippocampus system, see Ranganath and Rainer (2003) and Kumaran and Maguire (2007) for two reviews. Notwithstanding the heterogeneity of the different neural mechanisms promoting intrinsic motivations, they seem to have the same adaptive function: to drive the agent to acquire knowledge and competences without any extrinsic feedbacks. Such knowledge and skills can be exploited later, e.g. in adulthood, to attain biologically useful outcomes (Singh et al, 2010, Baldassarre, 2011).

Unfortunately, the study of the acquisition of new skills and knowledge was mostly carried out with experiments involving tasks either based on rewards (involving mainly animals, see Balleine and Dickinson (1998)), or on predefined task goals (involving mainly humans, see Elsner and Hommel (2004)). This has contributed to generate a large psychological and neuroscientific literature on decision making and on goal-directed behavior (see Balleine et al (2008), for a review). However, to the best of our knowledge, very little research has been carried out on the acquisition of new knowledge and skills based on intrinsic motivations. This work proposes a new experimental paradigm to study the effect of intrinsic motivations on action-outcome relation learning and action selection in absence of a reward or of a predefined task goal. The experiment, carried out on three- and four year old children, uses an experimental apparatus specifically designed to study intrinsic motivations and other processes with children (Taffoni et al, 2012a), monkeys (Taffoni et al, 2012b), and humanoid robots (Taffoni et al, 2013). The goal of this study is to verify if: *i)* intrinsic motivations toward the board may be triggered by the novelty and by the surprising features of the produced stimuli; *ii)* motivations may be kept alive in absence of an external reward or goal, by experiencing of action-outcome contingency alone; *iii)* action-outcome contingency may promote the acquisition of action-outcome relations; *iv)* learning may be split into sub-components related to some features of the action (i.e. *where* the action should be performed; *what* should be done).

2 Method

2.1 Participants

Since children seem to develop the capacity to link their actions with environmental changes (causal mapping) with age (Hickling and Wellman, 2001), two groups with different mean age were enrolled: 12 three-year-old children (36.7±0.8 months, mean ± Standard Deviation) and 12 four-year-old children (48.5±0.8 months). Subjects were recruited from a day-care center and were individually tested in a quiet and familiar room of the center. Parents of the children signed a written informed consent[1] describing the purpose of the experiment. The study involved tasks requiring free exploration of an experimental apparatus, which will be presented in the following section. Such kind of task requires a good level of attention span. A pilot experiment (Taffoni et al, 2012a) was carried out on twelve subjects aged between 24 and 68 months to test the equipment. We did not consider younger ages to avoid limitations stemming from lack of understanding of the task, or insufficient attention span, or insufficient motor coordination. Preliminary results of the pilot study led us to focus our investigation on three and four year old children. Indeed, children younger than three years of age were not able to keep their attention focused on the board for the necessary length time without the intervention of the experimenter, while children older than five found the task boring, so much that they did not perform it.

2.2 Stimuli and apparatus

A programmable apparatus was developed to investigate free exploration of the world, which is known to be a primary activity in children's motor knowledge and skill development. This apparatus, called mechatronic board, allows to control two key elements that facilitate free exploration: (i) the use of complex, unexpected, and surprising stimuli triggered by actions; (ii) the introduction of unknown causality links between those stimuli and the children's action (i.e. action-outcome relations to be discovered). It is composed of a planar base (WxHxD: 650x500x450 mm) and a frontal unit (WxHxD: 650x120x400 mm), see Fig. 1.A. The planar base is provided with three slots (180x180 mm) where different smart-objects (i.e. objects instrumented with sensors to measure the interaction with the user) can be plugged in. For this study, three simple round pushbuttons (diameter 60 mm) were used: a Blue Button (BB) on the left, a Red Button (RB) in the center, and a Green Button (GB) on the right. A unit for delivering both visual and acoustical stimuli was mounted above each button. The frontal unit contains three boxes, closed by sliding doors and controlled to open/close in a reprogrammable way by the actions performed on the buttons. A unit for delivering both visual and acoustical stimuli was mounted above each box also in the frontal unit.

2.3 Procedure

Each child was tested in a single session. Before any experimental session child and experimenters played together to familiarize. Subsequently, the experimenters showed the board to the child, saying that it was a *magic toy* for her/him. The experimental session started when the child sat on a chair in front of the board and was ready to begin the exploration. It was composed of three phases: Baseline, Learning, and Test, always presented in the same order. Subjects were randomly assigned to two different groups: EXPerimental (EXP) group and ConTRoL (CTRL) group. Each subject of the CTRL group was yoked with one subject of the EXP group, matched for age. The protocols administered to the two groups solely differed in the Learning phase.

The Baseline phase: This phase was the first to be administered. The goal of this phase was to estimate the initial skills of children and their interest in exploring the board. It lasted 5 minutes. During this phase, the audio-visual stimuli solely came from the planar base: whenever a button was pressed, the lights above it switched on and a xylophone sound was produced (three different tones corresponded to the three different buttons).

The Learning phase: during this phase, children were allowed to play with the board and to freely explore it. For the EXP subjects, the board was programmed to respond to any pressing of the buttons with contingent visual and auditory stimuli and to open a single box when its specific button was kept pressed for more than one second. A Simple Pusch (SP), i.e a button pressed for less than $1s$, switched on the lights above the button (on the planar base) and produced a xylophone sound as in the Baseline. An Extended Push (EP), i.e. when the button was pressed for more than $1s$, produced the same stimuli as a SP from the planar base, but it also produced the opening of a box (always empty in this phase) and the corresponding visual and audio stimuli from the frontal unit: the interior of the

[1] approved by the local Institute Ethical Committee of the Università Campus Bio-Medico di Roma, Prot. 10.CI_REV(05).12. ComEt-CBM 07/2012

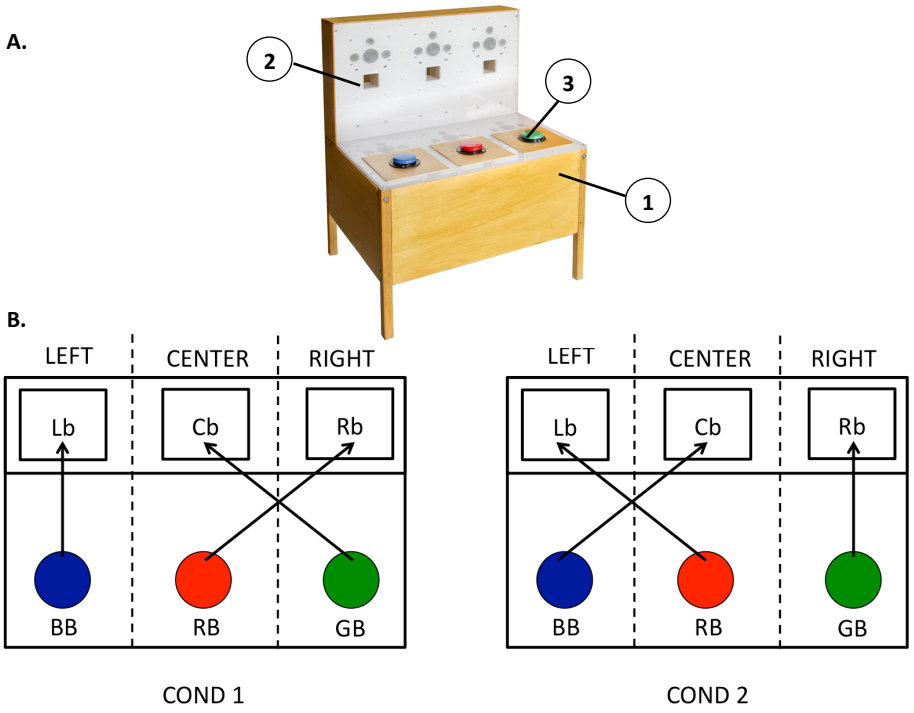

Fig. 1 A. The Mechatronic board. 1) the planar base; 2) the frontal unit: 3) mechatronic modules, in figure simple pushbuttons. B. Relations between buttons and boxes: COND 1) Crossed relations on the right side of the board; COND 2) Crossed relation on the left side of the board. Abbreviations: Lb, Left box; Cb, Central box; Rb, Right box; BB, Blue Button; RB, Red Button; GB, Green Button

box lit up, the lights above the box switched on, and the speaker near the box produced an animal sound (a different one for each box: a rooster's, a frog's or a cat's call). The relations between buttons and boxes was programmed to be direct (the button opens the box in front of it) or crossed (the button opens the box on its left or right side), see Fig. 1.B. Half of the subjects for each age group were tested with crossed relation on the right side of the board (Condition 1, COND 1) and the other half with crossed relation on the left side (Condition 2, COND 2). The CTRL children were yoked to the EXP ones: the mechatronic board recorded how the CTRL subjects interacted with it, but it was programmed to deliver the outcomes of the actions performed by their paired EXP subjects. In this way, CTRL subjects received the same number and kind of stimuli as their paired EXP subjects, but independently from their actions. This artifice prevented CTRL subjects from learning any action-outcome relationship. Moreover, it allowed to identify the different effects of action-outcome contingency and of unexpected events on children's behaviour. In particular, it allowed to verify how and how much these two aspects of the stimuli may promote or undermine the intrinsic motivation to interact with the board and if these mechanisms depend on age.

The Test phase: in this phase, a sticker was used as a reward, introducing an external goal to promote the child's actions. The outcomes depended on the subject's actions for both EXP and CTRL groups, so both of them could experience the action-outcome contingency. Relations between actions and outcomes were set to be the same as the ones proposed in the Learning phase to the EXP group, for both the CTRL and the EXP subjects. Each CTRL subject was tested using the same relations between buttons and boxes as his/her paired EXP subject. The Test phase consisted of 9 trials. During each trial, the subject was asked to retrieve a sticker placed in one of the three closed boxes (the sticker was always visible as the box door was transparent). Three different sequences of the sticker position were used[2] in order to avoid a bias effect due to the presentation order of the reward. The sequences were randomly assigned and counterbalanced among EXP subjects. Paired CTRL subjects received the same sequence

[2] **seq A**: Lb Cb Rb Lb Cb Rb Lb Cb Rb; **seq B:** Cb Rb Lb Cb Rb Lb Cb Rb Lb; **seq C:** Rb Lb Cb Rb Lb Cb Rb Lb Cb

Table 1 Experimental sample: for each subject the experimental setting used to program the board is reported as well as the reward sequences followed in the Test phase.

	EXP			*CTRL*		
AGE	code	cond	seq	code	cond	seq
3 YRS OLD	001	I	A	004	I	A
	002	I	B	005	I	B
	003	I	C	006	I	C
	007	II	A	010	II	A
	008	II	B	011	II	B
	009	II	C	012	II	C
4 YRS OLD	013	I	B	017	I	B
	014	I	C	018	I	C
	015	I	A	016	I	A
	019	II	A	022	II	A
	020	II	B	023	II	B
	021	II	C	024	II	C

order (Table 1). To open the box and retrieve the sticker children had to keep the correct button pressed for at least one second. Children were encouraged to retrieve the sticker without any suggestion on the action to perform. Each trial began with the reward inside a box and finished within two minutes, or earlier if the child took a shorter time to retrieve the reward. When the subject succeeded in opening the door and getting the reward, a new reward was placed inside the next box of the sequence. When the subject did not get the reward within two minutes, the same reward was moved inside the following box. The testing session ended when the nine trials were concluded, regardless of success in getting the rewards. The goal of this phase was to verify if children were able to exploit the skills acquired during the previous phase, in which only EXP subjects were allowed to understand the action-outcome relations.

2.4 Measures

The explorative behavior of each EXP and CTRL subject, during the Baseline and Learning phase, was assessed in terms of:

- Number of Pushes (NP): the total number of pushes, both SP and EP;
- Frequency of Pushes (FP): NP divided by the phase duration expressed in minutes (NP/min);
- Push Percentage (PP) of each button: NP of each specific button divided by the total NP.
- Number of Extended Pushes (NEP);
- Extended Push Frequency (EPF)
- Extended Push Percentage (EPP): the ratio between NEP and NP;
- Mean Holding Time (MHT): the mean value of the Holding Time (HT) of all the pushes, where the HT is the time (s) the child keeps the button pressed;
- Holding Time Standard Deviation (HTSD): the standard deviation of the HT of all the pushes;

During the Test phase, the exploration behavior and the performance of each EXP and CTRL subject were assessed separately for each trial, adding the following indexes to the ones described above:

- Spatial Correctness Index (SC): the difference between the number of Correct button Pushes (CP) and the number of Wrong button Pushes (WP), divided by the NP; its values belong to the range [-1, 1], where -1 indicates the complete absence of spatial correctness in the subject's pushes (he/she never pushes the right button), 0 indicates the subject performing an equal number of wrong and right pushes, and 1 indicates the subject understanding the Button-Box relation (she/he only presses the correct button);
- Time to Reward (TtR): the time (s) used by the subject to retrieve the reward (the value is considered as 'NaN' , i.e. Not a Number, if the reward was not retrieved);
- Trials to Criterion (TtC): the number of WP before the first Right Push (RP).

In addition, the Number of Rewards (NR) retrieved by each subject during the whole Test phase was measured and used as a subject's performance index. To select the most appropriate statistical test to be performed, the assumptions of normality and variance homogeneity of the relevant variables were verified each time. When worthwhile, appropriate data transformation (logarithmic, square root, or arcsin transformation) was used. In the case of assumption failure, despite transformation, non-parametric tests were performed, as reported in the next section.

3 Results

The initial skills of EXP and CTRL groups were assessed using the data collected during the Baseline phase. No significant differences were observed between the two groups. The modality of exploration seems to be affected by age: a repeated-measures ANOVA reveals a slight statistically significant difference in the PP of the three different buttons in three-year-old children ($F(2,22) = 3.48$, $p = 0.05$). The post-hoc tests (Bonferroni correction) showed a preference for the central button versus the left one. No statistically significant differences, on the contrary, resulted in four-year-old children. No additional differences were observed in the present phase. This finding may be explained by a poorer motor coordination (not fully developed) of younger

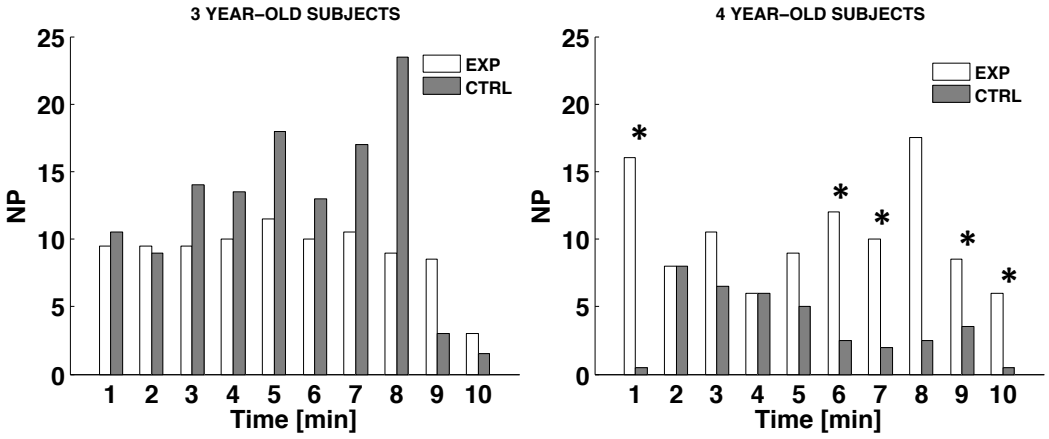

Fig. 2 The level of interaction with the board (expressed as NP) during the Learning phase: EXP (white bars) and CTRL (grey bars) subjects are compared, 3 year olds on the left and 4 year olds on the right. Four year-old CTRL subjects show a significantly lower level of interaction than the EXP ones in the first minute (Wilcoxon Rank-sum Test, p=0.04), as well as in the final part of the Learning phase (Wilcoxon Rank-sum Test: at minute 6, p=0.02; at minute 7, p=0.02; at minute 9, p=0.05; at minute 10, p=0.02;). On the contrary, 3 year-old CTRL subjects maintain their level of engagement with the task consistent with the one of the EXP group.

children, who tend to prefer exploring buttons which are simpler to reach.

During the Learning phase of the EXP subjects, the lights above the button remain turned on while children keep the button pressed. This feedback can facilitate the discovery of the effect of an EP and, subsequently, it may sustain the motivation to press the button. CTRL subjects cannot experience action-outcome contingency: they see the board turn on and off without any apparent reason. Such condition allows to clearly identify the different effects of action-outcome contingency and of unexpected events on children's behavior. We split the Learning phase into 10 time bins, each one lasting 1 minute. Subsequently, we measured the NP in each time bin to assess if children were involved in the task. The data of CTRL subjects were then compared to the EXP ones, considering three- and four-year-olds separately (see Fig. 2). While three-year-old CTRL subjects show a high level of interaction from the first minute, four-year-old CTRL subjects seem reluctant to explore the board in this first time interval. The involvement in the task of three-year-old CTRL subjects increases until the eighth minute when it reaches the highest value measured in all the experiment. After the eighth minute, the level of interaction drops. After the first minute, four-year-old CTRL subjects show a pattern of involvement consistent with the one observed in age matched EXP subjects until the fifth minute; then, differently from the three year-old CTRL subjects, their level of interaction decreases substantially, showing significant differences with the matched EXP subjects (at minutes 6, 7, 9 and 10, as Fig. 2 shows). This behavior is different from the one observed in 3 year-old subjects, where no statistically significant differences result in the level of interaction between EXP and CTRL subjects. The observed opposite pattern of interaction in the two CTRL groups, with younger and older children respectively increasing and decreasing their board exploration, may be due to an age dependent effect of action contingency and surprise/novelty in motivating them. Younger CTRL participants are strongly motivated to explore the board by the unexpected and novel events caused by the yoked condition: experiencing these events drives their exploration and keeps their interest for the board high (their level of interaction is similar, and even higher, to the one of EXP participants). On the contrary, four-year-old CTRL children seem to be motivated to explore the board by the possibility of learning a relation between actions and outcomes. During the first minute, four-year-old CTRL participants experience a high prediction error, as they see the board being activated without apparent reasons. This promotes their exploration from minute two to minute five. However, the impossibility of experiencing a causal relation between action and outcomes soon undermines the motivation to further explore the board (see De Charms (1968)). On the contrary, their coupled EXP participants maintain a high level of exploration as they have the possibility of experiencing a feedback coherent with their action. Such possibility

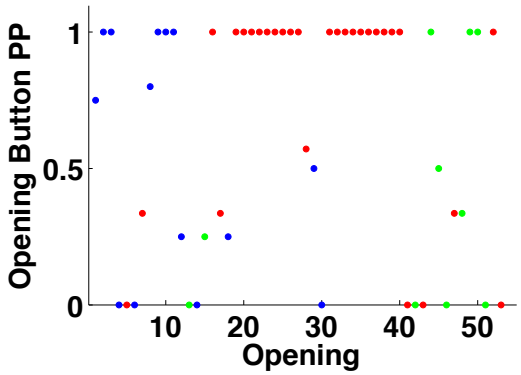

Fig. 3 Example of focusing behavior during the Learning phase in a single child (Subject 2). Each circular marker represents a Box Opening (BO): the color of each marker (blue, red, green) corresponds to the color of the button used to open the box (Opening Button). The y-axis shows the PP of the Opening Button, measured after the last BO and before the following one. A PP equals to one means a focalization on the button, which had caused the last BO.

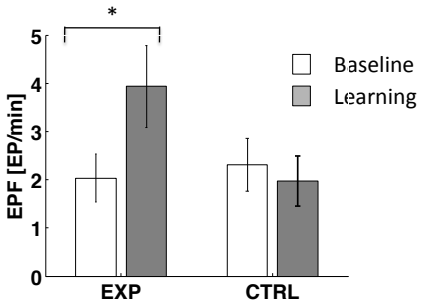

Fig. 4 Effect of novel unexpected stimuli experienced in the Learning phase: both three- and four-year-old EXP subjects, who had the possibility to experience these effects contingent with their actions, increased significantly their tendency to perform the action (EP) causing the unexpected effect (BO) with respect to the Baseline.

promotes their exploration and keeps their engagement with the task high. This finding seems to suggest that the purely novel and surprising aspects of the stimuli may be strongly motivating in three-year-old subjects, even in the absence of action contingency. On the contrary, in four year-old subjects, the action-outcome contingency seems to be the strongest motivating aspect, whose absence clearly dissuades from keeping the interaction.

The new experience of the box opening may promote a focusing behavior in EXP subjects: they might focus their actions on a specific button (spatial focusing) or they might repeat that particular interaction which caused the unexpected outcome, i.e. the EP, regardless of the button (action focusing). To investigate spatial focusing, the subject's actions after each BO were analyzed. No spatial focusing was observed, except for one EXP subject (Subject 002, 3 years old, COND 1). This subject explored each of the three buttons for a prolonged time and seemed to change the explored button after she/he experienced the opening of a different box. Subjects experiencing the outcomes after an EP should prefer this pushing modality to SP: we named this behavior action focusing. In order to investigate the action focusing, the EPF measured in the Baseline phase was compared with the EPF measured in the Learning phase. EXP children significantly increased the EPF in the Learning phase compared with the Baseline (Two-way mixed ANOVA, within-subjects effect, $F(1,10)=18.7$ $p < 0.01$), without any dependence on the age (Two-way mixed ANOVA, between-subjects effect, $F(1,10)=0.3$, $p > 0.05$), as shown in Fig. 4. This suggests that the experience of new and unexpected visual and acoustical stimuli (BO) contingent with the action (the EP) promoted its execution regardless of the button. No differences were observed for CTRL subjects, neither in three- nor in four-year-old subgroups.

During the Test phase, the effect of the different Learning phases of the two groups was assessed. According to the experimental design, during the Learning phase only EXP subjects could acquire knowledge about the functioning of the board, thanks to their free exploration of the device. In detail, two aspects of action selection were assessed: action learning, i.e. the understanding that an EP causes the BO; and spatial learning, i.e. the understanding of the right relation between the button and the box which is controlled by it. These aspects may be assessed using two metrics, respectively the EPP and the SC. If the subject has learnt that an EP opens the boxes, she/he is expected to perform it more frequently when asked to retrieve a sticker put inside one of the boxes. Similarly, if she/he previously understood the spatial relation between buttons and boxes, the correct button should be pressed more frequently than the others. A preliminary exploratory data analysis was carried out. The nine trials of the Test phase were grouped into three triplets, each one including one direct trial and two crossed ones. To assess what EXP and CTRL subjects learned in the Learning phase, and to verify whether they continued to learn during the Test phase or not, the evolution of mean SC vs. mean EPP was observed in the three triplets of the Test phase. In Fig. 5, the red markers represent CTRL subjects and the blue markers EXP subjects. Markers of coupled subjects have the same shape. The plot area is split into four rectangles by a vertical axis placed at

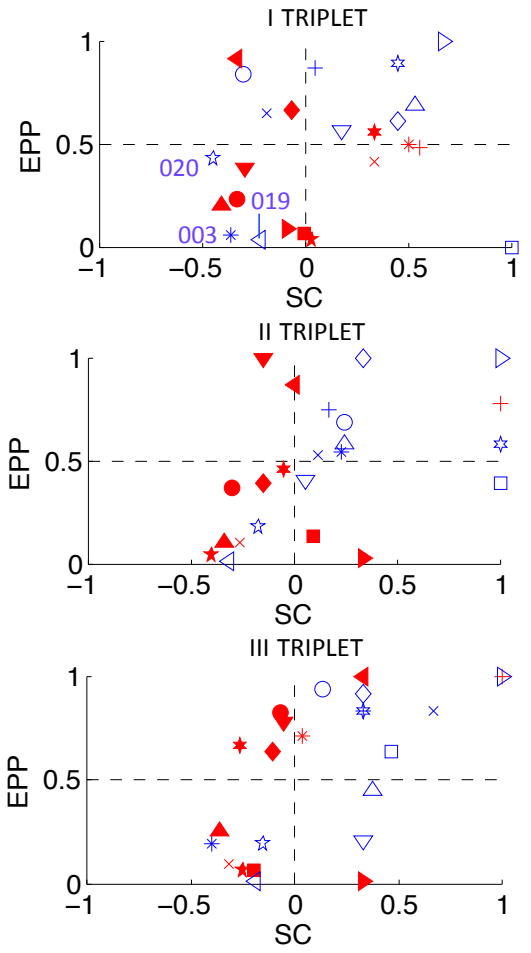

Fig. 5 Extended Push Percentage (EPP) vs. Spatial Correctness Index (SC) in the three triplets of the Test phase. The red filled markers are the CTRL subjects; the blue markers the EXP ones. When EPP is higher than 0.5, the subject prefers extended to single pushes. When SC is higher than zero, the number of correct pushes is higher than the number of wrong pushes.

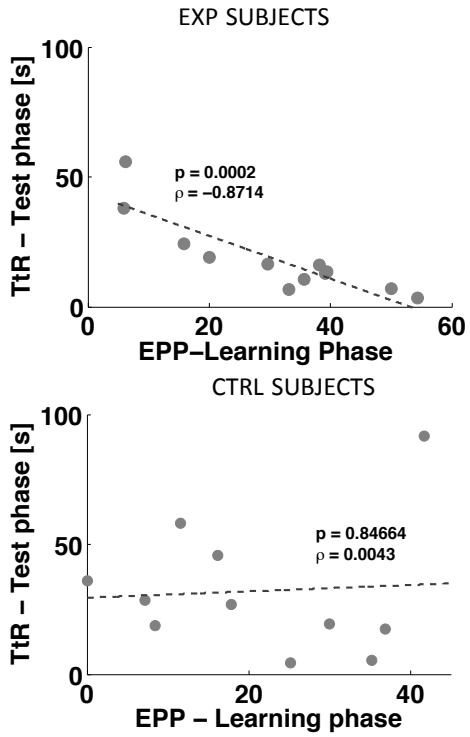

Fig. 6 Effect of action contingency on action learning: (top) EXP subjects who experienced the EP during the Learning phase more frequently, took less time to retrieve the reward during the Test phase. This relation is not present in CTRL subjects (bottom).

0 SC and a horizontal axis placed at 0.5 EPP, both indexes measured in the Test phase. These two levels were chosen since they represent two thresholds over which there is a focalization on the correct spatial relation and on the correct action, respectively. For this reason, the four rectangles represent four different levels of learning: no learning at all (left bottom rectangle); only action learning (upper left rectangle); only spatial learning (bottom right rectangle); both spatial and action learning (upper right rectangle). In the first triplet, 6 EXP subjects acquired both spatial and action knowledge versus only one CTRL subject, as expected. Only EXP subjects, in fact, had the possibility to learn action and spatial relations during the Learning phase. During the Test phase, EXP subjects seem to refine their SC: subjects with an SC higher than 0 increased from 7 in the first triplet to 9 in the last one. On the contrary, CTRL subjects seem to understand the action controlling the box opening: the number of subjects with EPP higher than 0.5 increased from 3 in the first triplet to 7 in the last one. The three EXP subjects, being in the no learning region in the first triplet, remained in this region in the last one. The reason for this finding may be that these subjects (003, 019 and 020) performed a lower NEP during the Learning phase and so they were not able to experience the correct action enough times.

The preliminary exploratory data analysis discussed, seems to suggest that action learning is acquired by EXP subjects during the Learning Phase. In particular, we hypothesized that the EXP subjects, who more frequently experienced the effect of EP during the Learning phase, were able to more efficiently retrieve the rewards in the Test phase. This hypothesis is confirmed by the correlation, proper of the EXP group, between TtR, measured in the Test phase, and EPP, measured in the Learning phase, as shown in Fig. 6. The higher the EPP in the Learning phase, the shorter the time needed by subjects to retrieve the reward in the Test

phase. The correlation is also significant when three- and four-year-old EXP subjects are considered separately (3-year-old subjects: p=0.01, R=0.81; 4-year-old subjects: $p < 0.01$, R=0.93), whereas it is not present at all in CTRL subjects, who were not allowed to learn any action-outcome relations during the Learning phase. Regarding spatial learning EXP subjects presented a significantly higher level of SC than CTRL subjects (Wilcoxon rank-sum test, Z=2.8, $p < 0.01$), as expected. To assess whether this knowledge had been already acquired during the Learning phase or it developed during the Test phase, the evolution of SC was analyzed in the three triplets of the Test phase. If a subject had already acquired spatial relations, this index is higher than zero. SC of the two groups is reported in Fig. 7 for each triplet. At the beginning of the Test phase, EXP subjects had not understood the spatial relations between boxes and buttons yet: in the first triplet, the difference between SC and zero only showed a small positive trend that was not statistically significant, meaning that the number of correct and wrong pushes was statistically the same. This index positively differed from zero starting from the second triplet in the EXP group, confirming that EXP subjects refined the spatial relation between buttons and boxes during the Test trials, and not during the Learning phase. However, the Learning phase is mandatory to prompt such learning process, which is indeed not observable in CTRL subjects. In this group the index was always statistically equal to zero, demonstrating that CTRL subjects, who were not allowed to learn anything during the Learning phase, did not have enough time to acquire the spatial relation during the nine trials of the Test phase.

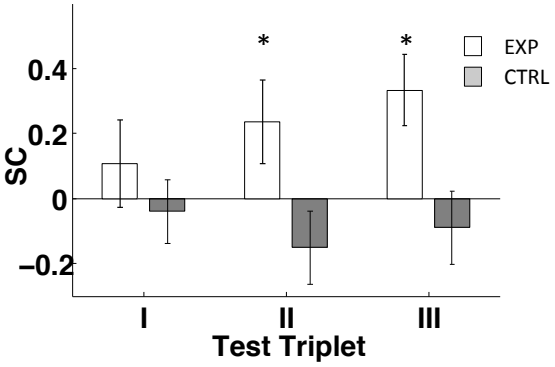

Fig. 7 Spatial learning during the Test phase: the SC is learnt by EXP subjects during the Test phase. The asterisk marks triplets where SC is significantly different from zero

Remarkably, when looking at the total amount of pushes, CTRL subjects revealed, in the Test phase, an explosion of interest for the board (paired t-test on FP, t(22)=-2.3, p=0.03), which was not observable in the EXP subjects (t(22)=-0.9, $p > 0.05$). Since CTRL and EXP subjects, along this phase, were equally exposed to the reward, the observed difference is not attributable to this mechanism. On the contrary, in the Test phase CTRL subjects began to experience BO, thus FP burst can be justified by the novel achievement of permanent action-outcome contingency.

Finally, the performance of EXP subjects in the Test phase resulted to be overall better than CRTL. Although EXP and CTRL subjects retrieved the same number of rewards (Wilcoxon rank-sum test, Z=1.7, $p > 0.5$), EXP subjects performed fewer pushes (Wilcoxons Rank-Sum Test on NP, $Mdn_{CTRL} = 10$, $Mdn_{EXP} = 3$, Z=-3.75, $p < 0.01$), needed less time (Wilcoxons Rank-Sum Test on TtR, $Mdn_{CTRL} = 11.2$, $Mdn_{EXP} = 7.2$, Z=2.2, p=0.03), and executed higher EPP (Wilcoxons Rank-Sum Test on NP, $Mdn_{CTRL} = 33\%$, $Mdn_{EXP} = 50\%$, Z=2.5, p=0.01).

4 Discussion

The present experiment consisted of allowing children to discover the relevant feedback stimuli of the mechatronic board, and to learn the action-outcome relations to recall a specific action when, at a later stage, the related outcome becomes desirable (i.e., it becomes an actively pursued goal).

The experiment presented in this work, even if on a small number of subjects, shows that free exploration of the world is sustained by the discovery of a hidden causal relation. Age has an effect, in this process. Children are able to develop an understanding of causal relations from external events at very early age, as shown by the fact that two-year-old children can express causal prediction (Hickling and Wellman, 2001). Even young infants seem to be able to infer some very basic laws related to the physics of the environment, as shown by the seminal work of Spelke et al (1992), or more recently by (Moll and Tomasello, 2010, Téglás et al, 2011, Mascalzoni et al, 2013). To refine this knowledge, children act like scientists who verify hypotheses (Gopnik et al, 1999). This enables them to gain a deeper knowledge of causal principles of everyday physics, biology and psychology by the age of five (Gopnik et al, 2004), clearly showing that this ability rapidly evolves with age. In our experiment, the EXP participants remained engaged in the task during the whole Learning phase, while the CTRL participants seemed to show an age related effect. In detail, 4-year-old CTRL participants were discouraged to explore the board by the apparent randomness of the stimuli, while 3-year-old ones were

promoted in their exploration. This finding seems to suggest that the purely novel and surprising aspects of the stimuli may be more motivating for three-year-old children than for four-year-olds. However, the absence of any kind of relation and the loss of the initial novelty and surprise cause a reduction in the interaction with the board also in these children. In fact, their level of engagement is comparable to that of 4 year-old CTRL children in the last part of the Learning phase, after the eighth minute.

The literature also proposes that actions are composed of different aspects, e.g. related to where the action is performed and what movements it involves (Redgrave and Gurney, 2006). The experiment described in this paper tried a preliminary investigation on the possibility that intrinsic motivations drive the learning of different aspects of actions in different ways. The possibility to experience action-outcome contingencies during the Learning phase guided EXP children's exploration: our findings suggest that children adopt an exploration strategy that allows them to learn the effects of specific actions (action knowledge) and the spatial relations (spatial knowledge) separately. Even if a focusing behavior similar to the one described in (Baldassarre et al, 2012) was not found during the exploration, action focusing related to the unexpected novel event was observed in those subjects who could experience the action-outcome contingency. A possible explanation of the absence of spatial focusing could reside in the effect of curiosity. Once the children experience a given button-evoked outcome, curiosity probably drives them towards the unknown effects of a different button. This indicates that novelty and contingency play a fundamental role in the action selection process.

Importantly, the experiment also showed that the knowledge acquired during free exploration, could improve the performance of subjects when they were subsequently asked to select the actions needed to accomplish useful, extrinsically-rewarding goals (gathering the stickers). In this respect, the experiment clearly showed that performance in these goal-directed tasks (e.g., the Time to Reward) was significantly correlated with having experienced the effects of their own actions (e.g. EPP) during free exploration. Our findings also suggest that action-outcomes contingency is not sufficient to acquire spatial relations when an external goal is missing: only by forcing a spatial focusing with the use of a reward (in our experiment a sticker), this knowledge has been acquired.

Finally, the experiment reported in this work allows investigation of different motivational effects of transient outcomes vs. permanent outcomes: lights and sounds may be considered as transient outcomes because they last for a short time contingent with the action; the box opening may be considered as a permanent outcome because it represents an environmental change lasting for a time sufficient to support further explorations and actions. To the best of our knowledge, this distinction is introduced here for the first time. Our findings seem to suggest that permanent outcomes have a higher motivational potential, as they more likely own a biological relevance, and also open up the possibility to perform further actions and explorations: both during free exploration, and with goal directed tasks, the perception of the impact of one's own actions on the environment proved to strongly enhance the interest towards the environment itself, in agreement with the proposals linking intrinsic motivations to competence (White, 1959) and mastery (Ryan and Deci, 2000).

5 Conclusions

Curiosity may be considered one of the driving forces that shape the process of acquisition of new skills and knowledge. The spontaneous exploration of the world plays a fundamental role in this process, providing subjects with an increasingly diverse set of opportunities for acquiring, practicing and refining new abilities. This activity is strongly promoted by the possibility to develop a causal mapping of the world. Children are motivated to explore the environment by novel events, which trigger their curiosity. Subsequently, their motivation to explore is kept high by the possibility to infer causal relations, thus increasing their knowledge and skills. This work studied how children develop the criteria for effective action selection. Three- and four-year-old children were observed during the free exploration of a new toy: half of them were allowed to develop the knowledge concerning its functioning, the other half were not allowed to learn anything. This study particularly focused on the way children acquire new knowledge during free exploration of the toy and how they reuse it for goal-directed tasks. The main results of this study can be summarized into four points: *i)* the novelty and the surprising features of the mechatronic board, in particular the transient effects that actions can cause on it (e.g., lights and sounds), triggered exploratory behaviors toward the board; *ii)* the purely novel and surprising aspect of the stimuli is far more exciting than the fact that the contingencies may allow a causal mapping for younger children; *iii)* action contingencies, and in particular outcomes lasting long enough to allow the performance of new actions and explorations (e.g., box opening), produced an additional motivating effect allowing subjects to acquire new knowledge on the possible outcomes that their actions caused on the board

and enhancing their capabilities to accomplish more effective, efficient, and desirable (extrinsic) goals; *iv)* actions are actually formed by sub-components (e.g., what, where, when, how) possibly relying on different cognitive/brain processes. Intrinsic motivations might lead to the acquisition of specific action components, at different times, within the overall learning process.

The afore mentioned points suggest that unexpected and surprising stimuli trigger children's exploration, even without a specific goal. This curiosity-triggered exploration is kept alive by the contingency between children's actions and the platform's outputs in four-year-old children, and by the joint effect of action contingency and novelty in three-year-old ones . In particular, action-outcome contingency allows the acquisition of new knowledge. By engaging the child for longer, it increases the likelihood of experiencing the effects of the action. When asked to perform a goal-directed task, children are able to apply the acquired know-how and to refine it. Future directions of this study may involve the electroencephalographic coregistration of the child's brain activity: monitoring the two subcomponents of the P300 wave (namely the P3a, related to the engagement of attention and the processing of novelty; and the P3b, related to unlikely action-related events (Polich, 2007)) may provide the neural correlate to understand if and when the action-outcome relations become known and predictable. Moreover, EEG recording, time locked to button pushes, may be used to investigate which brain areas are recruited to perform the task, how the hand controlling cortices of the two hemispheres interacts (Di Pino et al, 2012), and how learning processes may modulate sensorimotor integration (Ferreri et al, 2013).

Acknowledgements This work was funded by FP7-ICT program (project no. ICT-2007.3.2-231722 - IM-CLeVeR)

References

Baldassarre G (2011) What are intrinsic motivations? a biological perspective. In: Development and learning (icdl), 2011 ieee international conference on, IEEE, vol 2, pp 1–8

Baldassarre G, Mirolli M (2013) Intrinsically Motivated Learning in Natural and Artificial Systems. Springer

Baldassarre G, Mannella F, Fiore VG, Redgrave P, Gurney K, Mirolli M (2012) Intrinsically motivated action-outcome learning and goal-based action recall: a system-level bio-constrained computational model. Neural Networks

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37(4):407–419

Balleine BW, Daw ND, O'Doherty JP (2008) Multiple forms of value learning and the function of dopamine. Neuroeconomics: decision making and the brain pp 367–385

Baranes A, Oudeyer PY (2013) Active learning of inverse models with intrinsically motivated goal exploration in robots. Robotics and Autonomous Systems 61(1):49–73

Barto A, Mirolli M, Baldassarre G (2013) Novelty or surprise? Frontiers in psychology 4

Berlyne DE (1960) Conflict, arousal, and curiosity. McGraw-Hill Book Company

Botvinick MM, Niv Y, Barto AC (2009) Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. Cognition 113(3):262–280

Chentanez N, Barto AG, Singh SP (2004) Intrinsically motivated reinforcement learning. In: Advances in neural information processing systems, pp 1281–1288

De Charms R (1968) Personal causation: The internal affective determinants of behavior. Academic Press New York

Di Pino G, Porcaro C, Tombini M, Assenza G, Pellegrino G, Tecchio F, Rossini P (2012) A neurally-interfaced hand prosthesis tuned inter-hemispheric communication. Restorative neurology and neuroscience 30(5):407–418

Elsner B, Hommel B (2004) Contiguity and contingency in action-effect learning. Psychological research 68(2-3):138–154

Ferreri F, Ponzo D, Vollero L, Guerra A, Di Pino G, Petrichella S, Benvenuto A, Tombini M, Rossini L, Denaro L, et al (2013) Does an intraneural interface short-term implant for robotic hand control modulate sensorimotor cortical integration? an eeg-tms co-registration study on a human amputee. Restorative neurology and neuroscience

Gopnik A, Meltzoff A, Kuhl P (1999) The scientist in the crib: What early learning tells us about the mind. New York 1

Gopnik A, Glymour C, Sobel DM, Schulz LE, Kushnir T, Danks D (2004) A theory of causal learning in children: causal maps and bayes nets. Psychological review 111(1):3

Gottlieb J, Oudeyer PY, Lopes M, Baranes A (2013) Information-seeking, curiosity, and attention: computational and neural mechanisms. Trends in cognitive sciences

Hickling AK, Wellman HM (2001) The emergence of children's causal explanations and theories: Evidence from everyday conversation. Developmental Psychol-

ogy 37(5):668
von Hofsten C (2004) An action perspective on motor development. Trends in cognitive sciences 8(6):266–272
Kaplan F, Oudeyer PY (2007) In search of the neural circuits of intrinsic motivation. Frontiers in Neuroscience 1(1):225
Keen R (2011) The development of problem solving in young children: A critical cognitive skill. Annual review of psychology 62:1–21
Kenward B, Folke S, Holmberg J, Johansson A, Gredebäck G (2009) Goal directedness and decision making in infants. Developmental psychology 45(3):809
Kumaran D, Maguire EA (2007) Which computational mechanisms operate in the hippocampus during novelty detection? Hippocampus 17(9):735–748
Mascalzoni E, Regolin L, Vallortigara G, Simion F (2013) The cradle of causal reasoning: newborns preference for physical causality. Developmental science
Mirolli M, Baldassarre G (2013) Functions and mechanisms of intrinsic motivations: The knowledge versus competence distinction. Intrinsically Motivated Learning in Natural and Artificial Systems pp 49–72
Moll H, Tomasello M (2010) Infant cognition. Current biology 20(20):R872–R875
Nehmzow U, Gatsoulis Y, Kerr E, Condell J, Siddique N, McGinnity TM (2013) Novelty detection as an intrinsic motivation for cumulative learning robots. In: Intrinsically Motivated Learning in Natural and Artificial Systems, Springer, pp 185–207
Piaget J, Cook MT (1952) The origins of intelligence in children. WW Norton & Co
Polich J (2007) Updating p300: an integrative theory of p3a and p3b. Clinical neurophysiology 118(10):2128–2148
Ranganath C, Rainer G (2003) Neural mechanisms for detecting and remembering novel events. Nature Reviews Neuroscience 4(3):193–202
Redgrave P, Gurney K (2006) The short-latency dopamine signal: a role in discovering novel actions? Nature Reviews Neuroscience 7(12):967–975
Redgrave P, Vautrelle N, Reynolds J (2011) Functional properties of the basal ganglia's re-entrant loop architecture: selection and reinforcement. Neuroscience 198:138–151
Ryan RM, Deci EL (2000) Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology 25(1):54–67
Santucci VG, Baldassarre G, Mirolli M (2013) Which is the best intrinsic motivation signal for learning multiple skills? Frontiers in neurorobotics 7
Schembri M, Mirolli M, Baldassarre G (2007) Evolving internal reinforcers for an intrinsically motivated reinforcement-learning robot. In: Development and Learning, 2007. ICDL 2007. IEEE 6th International Conference on, IEEE, pp 282–287
Schmidhuber J (1991) Curious model-building control systems. In: Neural Networks, 1991. 1991 IEEE International Joint Conference on, IEEE, pp 1458–1463
Singh S, Lewis RL, Barto AG, Sorg J (2010) Intrinsically motivated reinforcement learning: An evolutionary perspective. Autonomous Mental Development, IEEE Transactions on 2(2):70–82
Smith L, Gasser M (2005) The development of embodied cognition: Six lessons from babies. Artificial life 11(1-2):13–29
Spelke ES, Breinlinger K, Macomber J, Jacobson K (1992) Origins of knowledge. Psychological review 99(4):605
Taffoni F, Formica D, Zompanti A, Mirolli M, Balsassarre G, Keller F, Guglielmelli E (2012a) A mechatronic platform for behavioral studies on infants. In: Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS & EMBS International Conference on, IEEE, pp 1874–1878
Taffoni F, Vespignani M, Formica D, Cavallo G, Di Sorrentino EP, Sabbatini G, Truppa V, Mirolli M, Baldassarre G, Visalberghi E, et al (2012b) A mechatronic platform for behavioral analysis on nonhuman primates. Journal of Integrative Neuroscience 11(01):87–101
Taffoni F, Formica D, Schiavone G, Scorcia M, Tomassetti A, di Sorrentino EP, Sabbatini G, Truppa V, Mannella F, Fiore V, et al (2013) The "mechatronic board": A tool to study intrinsic motivations in humans, monkeys, and humanoid robots. In: Intrinsically Motivated Learning in Natural and Artificial Systems, Springer, pp 411–432
Téglás E, Vul E, Girotto V, Gonzalez M, Tenenbaum JB, Bonatti LL (2011) Pure reasoning in 12-month-old infants as probabilistic inference. science 332(6033):1054–1059
White RW (1959) Motivation reconsidered: the concept of competence. Psychological review 66(5):297