D3.2 - Empirical Experiments on Intrinsic Motivations and Action Acquisition: Results, Evaluation, and Redefinition

Deliverable D3.2 of the EU-funded Integrated Project "IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots", contract n. FP7-ICT-IP-231722

USFD: T. Stafford, M. Thirkettle, K. Gurney, P. Redgrave
UCBM: D. Formica, G. Schiavone, F. Taffoni, F. Keller, E. Guglielmelli
CNR-ISTC-LOCEN: M., Mirolli, G., Baldassarre
CNR-ISTC-UCP: E. Polizzi di Sorrentino, V. Truppa, G. Sabbatini, F. Natale, E. Visalberghi

Lead Contractor for this Deliverable: USFD
Other contributing Beneficiaries: UCBM, CNR-ISTC
Deliverable due: 28 April 2011
Deliverable submission: 28 April 2011

Funding Institution: European Commission (European Union)
Project Officer: Cécile Huet
Funded under: Seventh Framework Programme
Work Programme Theme 3: ICT – Information and Communication Technology; Call identifier: FP7-ICT-2007-3; Challenge 2: Cognitive Systems, Interaction, and Robotics; Objective: ICT-2007.2.2 Cognitive System, Interaction, Robotics
Start: 01/01/2009 Termination: 30/04/2013

Cite this document as:

Stafford T. , Thirkettle M. , Gurney K., Redgrave P., Formica D. , Schiavone G. , Taffoni F., Keller F., Guglielmelli E., Mirolli M., Baldassarre G., Polizzi di Sorrentino E., Truppa V., Sabbatini G., Natale F., Visalberghi E. (2011). Empirical Experiments on Intrinsic Motivations and Action Acquisition: Results, Evaluation, and Redefinition. Deliverable D3.2 of the EU-funded Integrated Project "IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots", contract n. FP7-ICT-IP-231722.

Aims & Overview of the deliverable

Aims

The aims of the deliverable, as given in the original IM-CLEVER proposal were to identify new key empirical phenomena and processes, allowing the design of a second set of experiments.

This report covers:

(1) novelty detection and discovery of when/what/how of agency in experiments with humans ("joystick experiment") and Parkinson patients.

(2) how object properties that stimulate intrinsically motivated interaction and facilitate the acquisition of adaptive knowledge and skills in monkeys and children ("board experiment");

Overview

Experimental work on IM-CLEVER both informs the robotics and computational work which is conducted by other partners on the project, and is an arena in which ideas from other partners can be tested. The experimental paradigms developed for IM-CLEVER are all designed to work across species and subject groups (children, adults, patients).

So far initial development of the board and joystick tasks has been done, and we are now focussing on analysing the first generation of experimental results and planning new experiments. For details of these please see the sections covering the work at the three individual centres bellow.

The USFD work has focussed on the acquisition of novel actions. The work of UCBM and UCP has focussed on intrinsic motivation and cumulative learning.

Synergies between experiments and models

There is a close collaboration between computational and experimental work at USFD (only experimental work is reported here). Theoretical and modelling work by Prof. Gurney has informed focus in the experiments on two aspects of action learning which are of particular theoretical importance
1) Repetition bias.
2) Habituation of intrinsic rewards.

Ongoing modelling work with the joystick task in particular, promises to sharpen this focus in several directions:
3) The influence of exploration strategy on learning performance
4) The effect of partial knowledge (or lack of it) on learning performance.

Repetition bias (point 1 above) is a prediction of Redgrave & Gurney (2006) and subsequent computational investigations which suggest that a temporary increase in action likelihood following initial reinforcement may play an important role in acquiring a novel action. Experimental work suggests that the joystick task is highly suited to investigating the existence and nature of this

phenomenon. This idea is at the heart of the experimental programme and initial results from the joystick task look like confirming this phenomenon.

Habituation of intrinsic rewards (2) is the idea that the dopaminergic response to novel outcomes reduces as that outcome become predictable. This is observed experimentally in animals, but it recent modelling work at USFD suggests it has a computational imperative. Thus, we have shown that, under habituation of a learning signal, an agent in a simple reinforcement learning task (grid world) will learn optimal strategies without an explicit cost function driving such behaviour. Moreover, if such habituation does not occur, then the agent will continue to learn and develop suboptimal behaviour. We predict an experiment test in the joystick task, whereby if habituation to the reinforcement signal could be reduced (e.g. by varying the nature of the signal somehow, or by pharmacological manipulation) then subjects would suboptimally widen their locus of search for the target 'hotspot' in a 'where' task after initially learning a close focus for this spot.

Input of experiments to the joystick task model (Task 5.2).

In addition USFD are developing biologically plausible models of the joystick task based on loops through basal ganglia controlled by topographically organised cortical regions responsible for reaching movements. This work is preliminary but the architecture will enable investigation of points 3 and 4 above. Thus, (in regards to 3) what happens in a 'where' task (say) if the subject adopts different strategies for exploration? Possibilities include a regular 'raster scan' of the reach space, a progressive search along radials from the centre of the space, spiralling away from this centre, etc. We anticipate differences in performance depending on strategy and our models will be able to address these issues because they are endowed with the topographic representation of reach.

In regards to point 4, we aim to extend our models to allow representation of 'gestures' in the 'what' variation of the task. Experimentally, the extension of the task to include this variation opens up several interesting possibilities about the use of partial task knowledge. Thus, we expect different search strategies and learning performance if subjects are told whether gesture and/ or location are relevant. These will all be able to be modelled within our framework.

The USFD theories of dopaminergic function in the basal ganglia, and the role of the whole network in allowing the learning of novel actions, are being incorporated into the models of CNR and, thus, in turn informing the development of the CLEVER-B2 architecture.

Input of experiments to the board experiment model (Task 6.1) and CLEVER-B2 Demonstrator (Task 7.4), and viceversa.

The design of CLEVER-B architecture by CNR-ISTC-LOCEN, USFD, and AU, has directly informed the experiments by CNR-ISTC-UCP with monkeys, and by UCBM with children, under two important respects:

1. The whole experimental paradigm was designed to investigate how monkeys and children *learn a repertoire of actions based on intrinsic motivations*.
2. The general structure of the experiment was suggested by the idea, derived from the computational theories and models on intrinsically-motivated cumulative learning, for which in an ecology where multiple things (skills and action-outcomes) can be learned, intrinsic motivations lead organisms to *focus learning* on one thing at a time in a cumulative fashion. This aspect is captured in the initial training phase of the experiment with monkeys and children, when the manipulanda are presented *all together* so to study how habituation for intrinsically interesting items can lead to explore all action-outcomes contingencies in sequence. Although we could have trained each button-box contingency separately, this

would have not allowed us to investigate the issue of habituation and learning of multiple affordances (skills and action-outcomes) in a single situation.

This focus has made the experiments both novel and closely linked to the core interests and computational models of the rest of IM-CLEVER.

The results on experiments themselves are giving important feedbacks and insights to the modelling activities. Although preliminary (the broad implications of experiments are still under scrutiny), these feedbacks and insights can be summarised as follows (further details will be given in other deliverables related to models and CLEVER-B2 demonstrator):

1. The exploration and learning of the environment affordances based on intrinsic motivations likely interact with the effects caused by extrinsic motivations and/or expectations of extrinsic rewards. See the final considerations related to the experiment with monkeys in this document on this. This interaction is very important for organisms and poses interesting computational challenges: what are the best mechanisms that can be used to regulate the relation between intrinsically and extrinsically motivated learning?
2. The experimental results indicate that the exploration of multiple affordances of a rich environment is a big challenge for organisms (and so for experiments as they create a strong variability). On one side, the results obtained so far seem to suggest that attention, and the capacity to focus it on one exploratory activity, play a key role in such environments. In relation to this, the experiments obtained so far might indicate different attentional capabilities between children and monkeys. Moreover, the videos of the experiments seem to suggest that in complex set-ups there might be smaller focussing and more uniform exploration than expected (unless there are dependencies between the different things to be learned).
3. Experiments with children and monkeys are suggested an insight that might be crucial for modelling. Action learning based on intrinsic motivations can be carried out based on two different processes: (a) bottom-up process: exploration lets organisms discover an interesting outcome, and this is progressively associated to the performed action that is progressively refined (USFD theory); (b) top-down process: exploration lets organisms discover interesting outcomes; when this happens these outcomes are stored in memory, activated as desired (so they become "goals"), and then used to drive the learning of actions necessary to accomplish them. This process is so a "goal-based action learning" process: CNR-ISTC-LOCEN is now developing this idea and comparing it with the process originally proposed by USFD. It might be that both mechanisms coexist in organisms, but that only more sophisticated ones (e.g., primates, humans) posses the second one in a substantial degree. This issue is becoming central in the debate related to CLEVER-B integrated model.

The information from experiments to computational models is expected to become even richer with further evaluations and reasoning on the experimental results.

Introduction the experiments on the discovery of Novel Actions ("Joystick task", USFD)

A versatile agent must learn new actions - options which cause specific effects in the world. We call this the action discovery problem.

The agency hypothesis (Redgrave and Gurney, 2006; Redgrave et al., 2008) proposes that the basal ganglia appear ideally configured to determine whether the agent is the likely cause of the unpredicted event. It proposed that dopamine(DA)-related repetition and neural plasticity underly the discovery of the causal com-ponents of behavioural output and that these components are encoded as novel actions. This proposal is based on an analysis of the functional architecture of the basal ganglia and considerations of signal timing. (Note we use the general term `agent' to accommodate bio-mimetic architectures used to control robot be-haviour (Prescott et al., 2006) as well as the neural systems in the brains of ani-mals that control their behaviour). For example, the relatively invariant timing of the phasic DA response (Schultz, 1998) highlights the importance of considering the signals that are also likely to be present in targeted structures at the time of phasic DA release, because it is with these signals that DA is most likely to inter-act.

In order to conduct behavioural experiments which can reveal aspects of action discovery, it has been necessary to develop a new experimental paradigm in which action acquisition can best be investigated. Such a paradigm requires that the learning agent discovers what aspects of its behaviour are responsible for evoking a salient sensory event. We have devised several versions of a joy-stick task which satisfy this requirement. Briefly, advantages of this task are that it can be performed by a wide range of species from mouse to man, also by robots. It can be used to study independently the different aspects of action acquisition, including 'where' a response has to be made, 'what' the response should be, 'when' it should be performed also 'how fast' the movement should be.

A second important feature of the joystick task is that it can support repeated measures experimental designs. For example, when the agent has dis-covered that a movement made to a specific location, or forming a particular gesture, or any movement made at a particular time or speed is causal in eliciting a sensory reinforcer, the criteria for reinforcement can be changed, thereby requiring novel actions to be discovered. This can be done repeatedly and will be important for investigating brain mechanisms contributing to action discovery. The option of determining performance before and after experimental treatments in the same subject will add considerable power to the experimental designs. Another important feature of the joy-stick task is that each of the 'reinforcement criteria' – the features of action that elicit a reinforcing signal – can be continuously varied in terms of difficulty.

The main experimental question to be addressed by USFD will be to investigate the sources of sensory input which can reinforce the acquisition of novel actions. This issue is particularly important because it will link with biological studies implicating the sensory-evoked phasic response of dopaminergic neurones in behavioural reinforcement. The specific question we are

addressing is whether the comparatively crude but fast sub-cortical sensory processing which we know can elicit phasic dopamine responses is solely capable of reinforcing novel action acquisition, or whether this can be supplemented by early sensory processing from the cerebral cortex. This is important because a range of more sophisticated perceptual competencies afforded by cortical sensory net-works would be capable of acting as reinforcers. These questions are being addressed, first, by using reinforcing stimuli that can be discriminated only by cortical sensory processing, e.g. determinations of colour and high spatial frequencies (to which sub-cortical visual systems are blind), and second by the use of masking paradigms which block conscious cortical analysis of sensory events while leaving subcortical sensory processing intact.

More details on the theoretical background to the USFD experiments can be found in the forthcoming book chapters (Stafford et al, 2010; Redgrave et al.2010; Gurney et al, 2010).

Joystick experiment set-up and protocols

The essence of the task is that the subject's free movements are recorded, either via a manipulandum such as a joystick, or directly via a touchscreen. Certain movements, henceforth ` targets', result in a sign or signal, henceforth the `reinforcement signal'. The aim of the task is to discover what characteristics of movement of the manipuladum evoke a phasic stimulus (which may or may not be a reward). The target may be defined in terms of a range of movement qualities: e.g. absolute spatial position, in which case it is a `hotspot', or in terms of a relative motion anywhere in absolute space, in which case it is a `gesture'. The target can even be related to the timing of the movement, e.g. onset or speed, regardless of it's spatial characteristics. The success of many real-life actions can depend on all of these components, which refer to as the `where', `when' `what' and ‘how’ qualities of action. For different experiments with the task the target will be defined in terms of one or more of these dimensions, so it is possible to investigate the discovery of different components of an action independently of the others. When one target has been learnt, the criteria for reinforcement is simply changed and a new action has to be discovered. This therefore affords the requirements of repeated measures.

Fig. 1. Experimental set-up for (a) humans and (b) rats, showing (1) manipulandum, (2) visual signal of reinforcement, (3) participant engaged in task and (4) food hooper for delivery of rewards to maintain behaviour (not shown for human subject).

Figure 1 shows the apparatus for running the experiment with both human and rat participants. Note that in the human set up the computer display is used only to deliver signals that the target motion has been made, it provides no visual feedback on the recorded position of the joystick. For the rat version, a long-handled manipulandum hangs from the ceiling of the rat's enclosure, to give it su±cient mechanical advantage. It can be moved with precision by the animal using a mouth or forepaw grip, or less precisely using a full body or tail swipe. Once moved, the rat joystick is engineered so that it maintains position rather than returning to the center point. While a typical computer-literate human participant can be simply instructed to make exploratory motions with the joystick, rat participants require more direction. For the rat versions of the task, so far, we have shaped the animal's behaviour by initially reinforcing any movement of the joystick and subsequently refining the required target. A full description of the mechanics and procedures involved in running this task with rats is in preparation. Similarly, we are also preparing a full description of the procedures involved in the human version of the task.

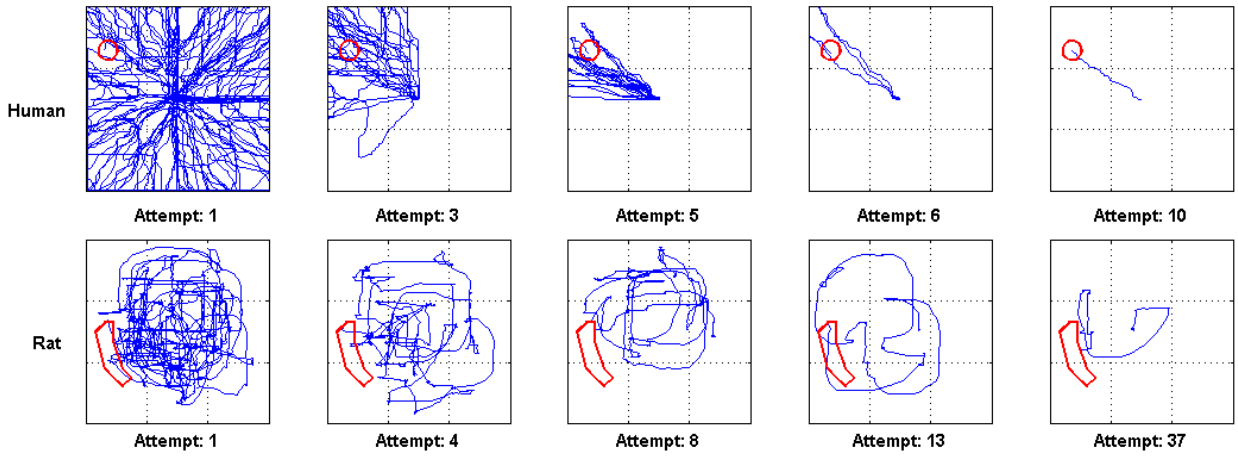


Fig. 2. Movement traces (blue) for a spatial target (outlined in red) for typical (a) human and (b) rat participants.

As alluded to, the task affords a complete record of the movements of the manipulandum. Figure 2 show typical continuous traces from both human and rat subjects as they initially explore and then refine their movements so as to home in on a spatially defined target (a spatial `hotspot)'. Note the similarity in the plots. Although rats take longer to refine movements into a novel stereotyped action, the similarity in the progression of behaviour in this `where' version of the task suggests that we are tapping the same process that relies on similar underlying machinery of action-discovery. Given the degree to which the basal ganglia have been conserved in mammalian evolution, this is to be expected. From the raw, total, data of participant movements various statistics can be computed which reveal the Progress of action learning. Figure 3 shows three sample statistics which all reveal the progress of action learning during a trial with a typical human participant, where `trial' refers to all attempts at finding a particular target. Figure 4 shows typical learning statistics from a rat. Note that within trial learning is evidenced, but no obvious across-trial learning.

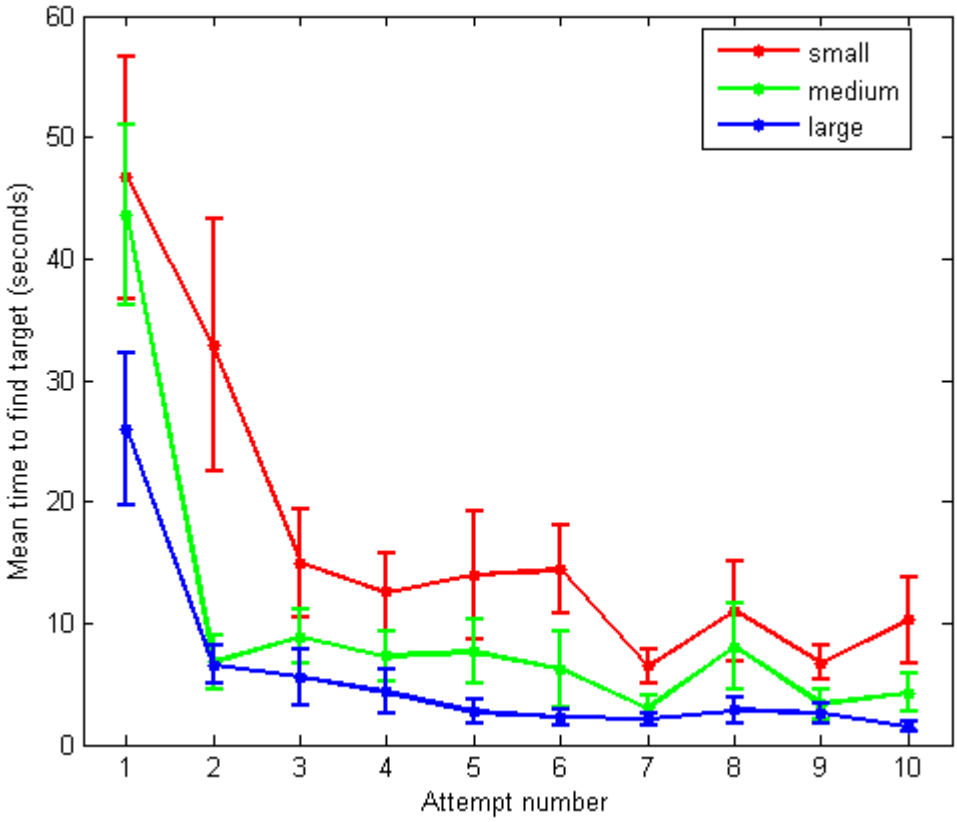


Fig. 3. Statistics showing within{trial learning for typical human participants (n=29, standard error bars shown

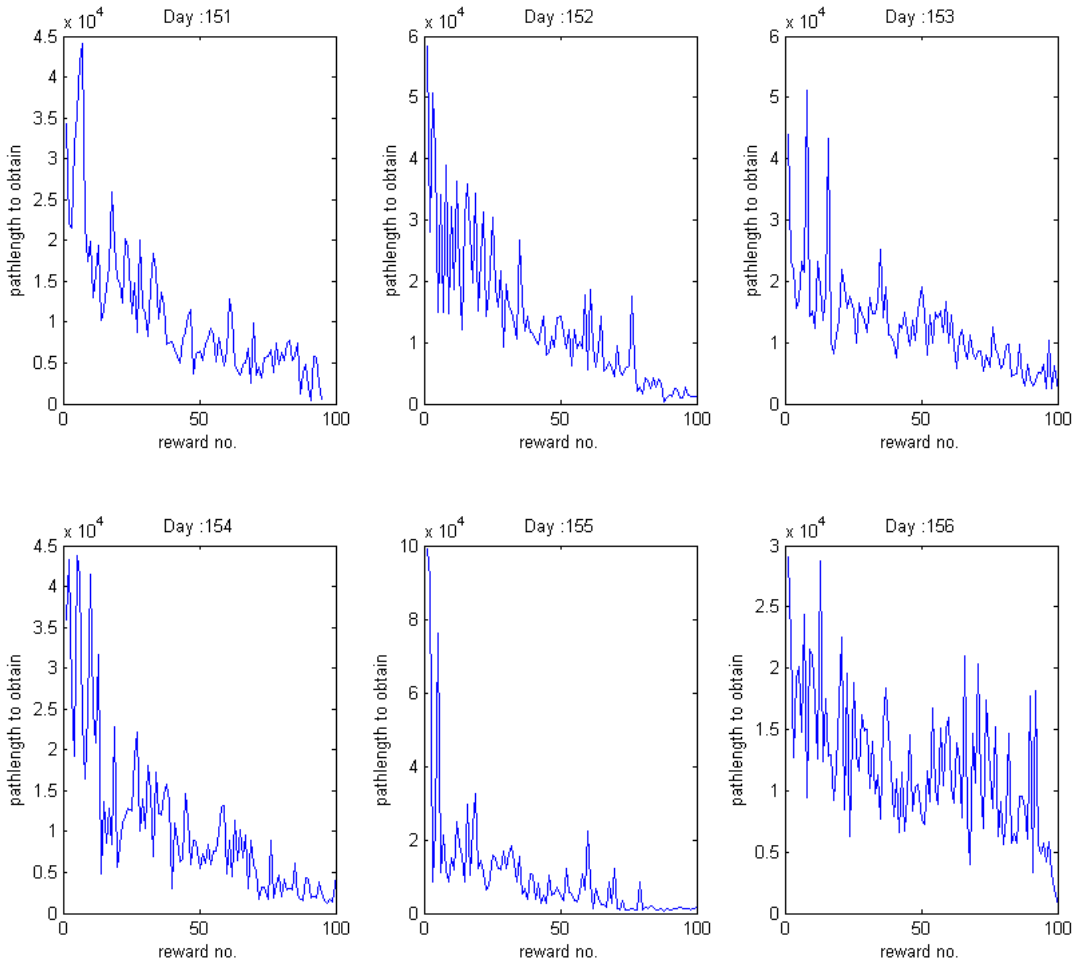


Fig. 4. Within-trial learning for a typical rat

Joystick experiment results

Firstly, the task development work described above has put us in the position where we are now able to test the specific predictions made in the original IM-CLEVER proposal, as well as opening up new possibilities for investigation (discussed below).

Secondly, we have completed the first of the specific experiments which we committed to in the original IM-CLEVER proposal.

Hypothesis & Overview

The hypothesis stems directly from our proposal that the neural substrate of this action-outcome learning exploits phasic signals from dopaminergic (DA) neurons within the basal ganglia (Redgrave & gurney, 2006), and from the fact that neuroanatomical evidence suggests that phasic DA signals are triggered by direct projections from the superior colliculus (SC). We therefore hypothesised that stimuli available to visual cortex but unavailable to the SC may be ineffective or less effective in supporting action acquisition.

We tested this using the joy-stick learning task and reinforcing signals defined by either collicularly available luminance information, or a cortically available signal solely detectable by the short-wavelength cone photoreceptors and therefore invisible to the collicular pathway.

The surprising result was that participants were able to learn effectively using reinforcing signals unavailable to the SC. Equivalent numbers of reinforcing signals were triggered in the collicular and cortical conditions, however, action acquisition was slower, demonstrating that learning of actions was less efficient, with the short wavelength stimuli. We conclude that luminance signals via SC are the most efficient, but not exclusive, reinforcer of novel actions.

Validating reinforcing signal types

Theoretical background to calibrated signal validation: Signals defined solely by colour discriminable only by the short-wavelength cone have been shown to produce slower reaction times than signals defined by luminance (Bompas, Aline, & Sumner. 2008) This amounted to a 23ms difference in manual reaction time. We replicated the experiment to verify that our calibration procedure was successful, and that the changes we had made in the display (the size of our signal on screen was greater than the original, but while the original experiment used a 250msec display, we used 12msec) had not altered the processing of the stimuli.

Procedure: This replication was conducted at the end of the calibration stage of the experiment, so that the data comes from the participants of the eventual joystick task. Participants had to decide whether a 12ms flashed stimulus occurred on the left or the right of the screen in a given trial. There were 80 presentations of Luminance and S-cone stimuli, and the experiment was run twice for each participant – once at the end of the calibration session, and once at the beginning of the joystick session which usually occurred a few days after the calibration session. Each participant produced 160 responses to luminance and s-cone stimuli, anticipatory responses were removed from the data set and the remaining correct responses were analysed.

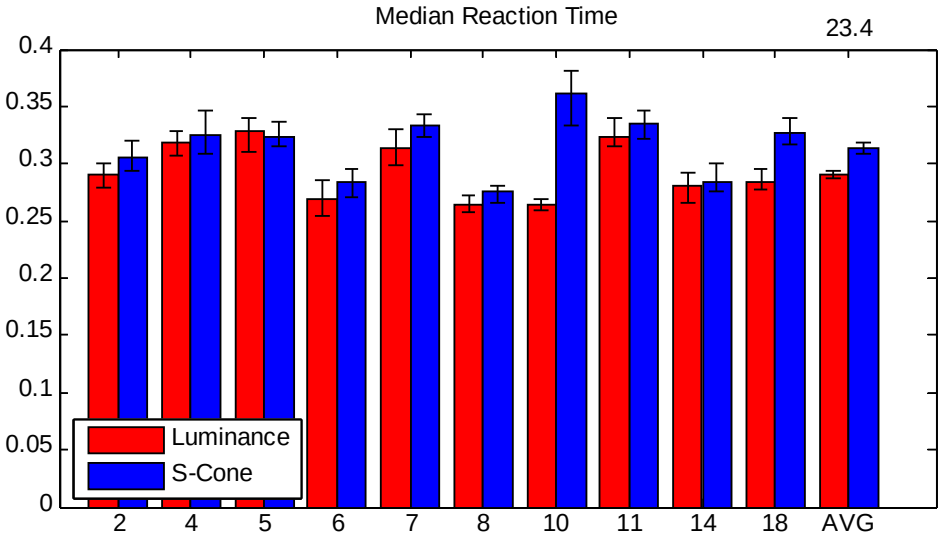
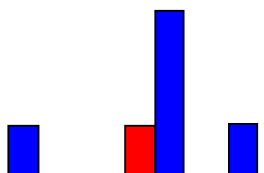


Figure 5: Median reaction time (seconds) for participants tested so far (participant numbers arbitrary)



Results: A 23.4 msec difference was found between s-cone and luminance signal RT, almost identical to the original finding so we can be confident that our calibration procedure is successful, and the changes we made have not had any untoward impacts upon perceptual processing. The reaction time difference is evidence that the two stimuli are being handled by different processing paths, we can now use this to study the reliance action-outcome learning mechanisms have on one of these processing paths: the luminance defined, collicular pathway.

A significant difference in action acquisition between cortically and subcortically mediated reinforcing signals

Theoretical background: Two hypothesis being tested. Chromatic information, unlike luminance, is processed in such a way that it has no direct connection to the basal ganglia via the superior colliculus. Therefore reinforcing signals defined solely by colour will either not produce learning, or learning will be impaired due to a processing delay. Any effect on learning the delay inherent to colour processing has should be replicable with a luminance signal by introducing an artificial delay between the successfully performing the action and presentation of the reinforcing signal.

Test procedure: In this form of the joystick task the participant blindly moves the joystick until they encounter the hotspot, then they receive a 12ms reinforcing signal. The hotspot is placed in a random location for each trial. The participant then uses this signal to guide their movements to reoccupy the hotspot frequently enough to denote success. Success is defined by 15 encounters of the hotspot within 1 second. Reinforcing signals are separated by a 30ms "refactory period" to avoid signals fusing together. The luminance and colour signal trials are blocked to prevent any interaction in the calibration or confusion in the participant. 9 temporal delay levels are used to map the impairment on learning 0, 75, 150, 225, 300, and 375 msec. Each delay level is repeated 3 times for each signal type to allow the average to be taken. One break is given during the task.
10 undergraduate participants completed the experiment using stimuli individually calibrated to them.

Data analysis: The location of the joystick is polled every 1msec and this becomes our dataset for analysis. The critical factor is the period between the participant's first encounter with the hotspot and success in the trial – the homing period. This can be considered in terms of:

- The duration of the homing period
- The distance covered by the participant during the homing period.
- Number of encounters with the hotspot before they achieved the 15 signals/sec necessary to end the trial.

Data is collapsed by signal delay and type to provide average duration, distance and number of signals presented.

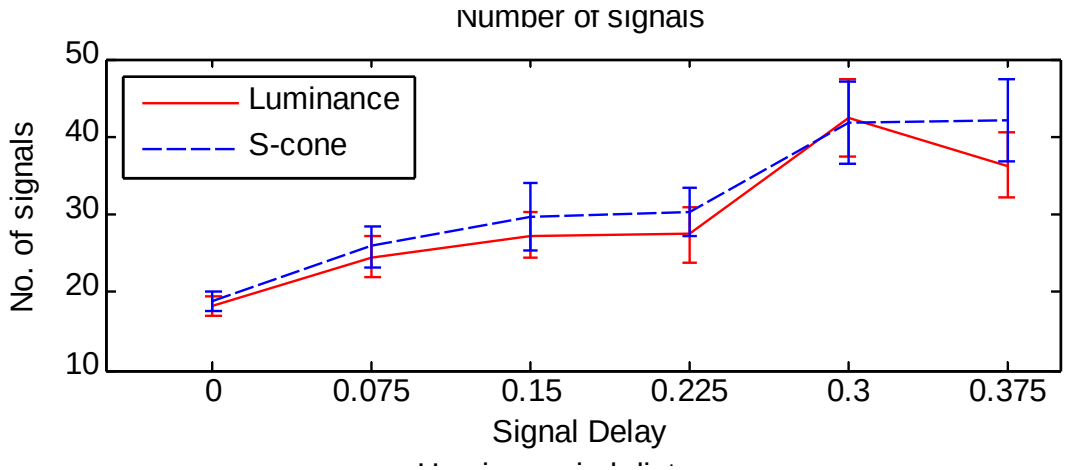


Figure 6: Reinforcement signal delay (seconds) against number of average number of signals gathered by participants. No significant differences between reinforcement signal conditions, suggesting that participants are receiving the same *amount* of reinforcement in each condition.

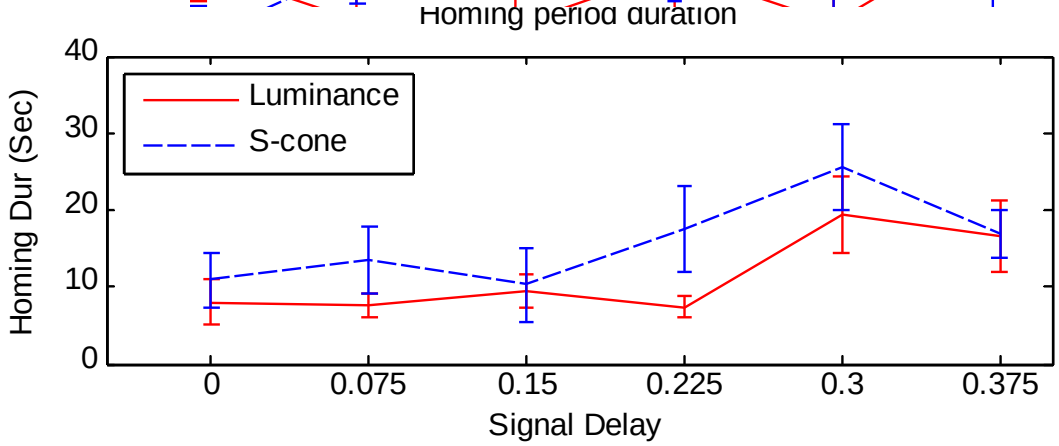


Figure 7: Reinforcement signal delay (seconds) against average time to find target. Learning based on the s-cone reinforcement signal is significantly impaired relative to learning based on the luminance signal.

Results: The expected effect of delay on learning performance was found – when a delay was placed between action performance and feedback, participants performed more poorly, learning slower. There was no difference between the number of signals needed (i.e. encounters with the hotspot before successful trial completion) when the feedback signal was directly available to the colliculus compared to when it was unavailable. Although no more signal presentations were required, participants were slower to use the feedback information to guide their behaviour and discover the hotspot on trials when the feedback information was collicularly unavailable. This demonstrates that while not the exclusive source of information to the action-learning mechanisms in the basal ganglia, the superior colliculus is the preferred supply of visual information. Learning based upon visual signals unable to take this processing route is impaired due to the delay imposed by the longer processing path taken by the information to the basal ganglia.

New experiments with the joystick task

The primary aim of the experiments was to test theories of the neurobiology of action learning, using different kinds of stimuli and Parkinson's Disease patients. Next we will
- test whether action learning is supported by subcortical or cortical routes, using stimuli manipulation (including 'visual masking' to make stimuli consciously invisible)
- test how Parkinson's Disease impacts on action learning.
These are our commitments as features in the original IM-CLEVER schedule of work.

This work has given us vital insight into the task, and allowed us to refine versions which reveal the high sensitivity of action learning to delays between action and effect (learning in this task impaired with delays of as little as 50 ms).

Another aspect of action learning that has become clearer is the different stages of action learning: exploration, action discovery and action refinement. In other words

- first you explore the space, generally
- then, after you trigger some reward, you focus your efforts on finding out what causes that outcome
- then, after you have identified the rough action which causes an outcome you refine that outcome and the action becomes smoother, faster and more automatic.

We hope to use the same task to give us insight into all these elements of action learning

We are also considering two new strands of investigation, both of which build on the existing IM-CLEVER work and could make significant contributions to the overall scientific aims of the project.

Before beginning our experiments we thought it prudent to conduct an in depth review of action control in Parkinson's disease (published in Nature Reviews Neuroscience; Redgrave et al, 2010, Framework 7 funding acknowledged). As a consequence of this review new information came to light and new hypotheses were proposed. However, one thing to become clear was that the original rational for the experiments with Parkinson's disease patients proposed in the original application can no longer be applied. The original assumption was that Parkinson's disease is associated with a non-selective degeneration of midbrain dopamine neurones and that testing subjects while on medication can distinguish between tonic and phasic activity of dopaminergic input to the basal ganglia. Following extensive collaborative exchanges with our clinical colleagues on this paper we now know that the loss of dopamine from the basal ganglia, for most of the time the patients suffer the disease is highly selective. The extensive loss differentially affects territories of the basal ganglia associated with the performance of habitual behaviour….which is probably not engaged with the initial acquisition of a novel action. Consequently, there would be little point in proceeding with the originally proposed experiments. However, in the light of our initial findings with the joy-stick task with different sensory reinforcers, we are proposing to revise our experimental programme to explore in more detail the what in which different aspects of action acquisition (where, what, when and how) can be separately acquired and the ways in which they may be integrated. It is our opinion that these investigations will better serve the overall aims of the IM-CLEVER project.

Stand 1: Habitual and Goal-direct control of action learning

It has been established that the same action (e.g. lever press) can be initiated either by goal-directed (outcome) or habitual (stimulus) control systems (Dickinson, 1980). Moreover, it has been shown that these two different modes of control engage different parts of the looped architecture that pass

through the basal ganglia (Balleine & O'Doherty, 2009). To study the reinforcement processes associated with basal ganglia circuitry it will be important to know: (i) whether different aspects of the joy-stick task engage different loops through the basal ganglia; and (ii) whether goal-directed and habitual control of each of the different aspects of action (what, where, when etc), also engage different functional territories in the basal ganglia, as suggested by the work of Yin, Knowlton and Balleine (2006). Most tasks are acquired initially by goal-directed control and become habitual only after many repetitions. Therefore, to discriminate goal-directed and habitual control, it will be necessary do develop versions of the joy-stick task where the different aspects of action acquisition can be subjected to the formal test of outcome devaluation. These procedures will be implemented as follows. When different aspects (where, what, when…) of an action are being discovered, subjects will have the opportunity of discovering two actions in the same trial, one that causes reinforcer A (e.g. food) and the other that causes reinforcer B (e.g. drink). Comparisons will be made between sessions where subjects experience regular shifts of the target criterion (e.g. the area that will elicit the reinforcer), and where subjects have experienced numerous trial sessions where the target criterion for reinforcement remains the same. In both cases, the outcome devaluation test will be achieved by permitting subjects prior access to one of the outcomes. If the behaviour is goal-directed there is a performance decrement on the devalued outcome, if it is habitual performance is maintained. The novel aspect of these procedures is that the different aspects of action acquisition will be subjected to analyses for goal-directed versus habitual control, as a necessary prelude to investigations of territorial engagement within the basal ganglia.

Strand 2: Cortical contributions to action learning assess using investigation of task performance in cortically blind monkeys

A unique opportunity has arisen where the aims of the IM-CLEVER project can be advanced by a novel collaboration with a Japanese group (lead by Profs Tadashi Isa and Atsushi Nambu) which was unforeseen at the time of the original IM-CLEVER proposal.

> *Note: this collaboration would be in synergy with the IM-CLEVER research, but would not be funded by any IM-CLEVER monies.*

One of us (Prof. Redgrave) has been invited to go to Japan as a visiting Professor to the National Institute for Physiological Sciences (NIPS) in Okazaki, to participate in a three month collaborative research project later this year (2011). Although this visit will be funded almost entirely by NIPS the envisaged collaborative project has been designed to have direct relevance to USFD's specific aims for the IM-CLEVER project. The relevance of the project derives from the fact that our Japanese collaborators have a unique population of monkeys that have sustained selective lesions of visual cortex. So far, the experiments conducted by our collaborators in Japan have tested the effects of V1 lesions, and have better defined residual visual competencies that can be attributed to the superior colliculus. All these experiments have tested the occulomotor performance of the lesioned animals. On the other hand, for the past few years we in Sheffield have been promoting the idea that visual stimuli in the colliculus can not only drive gaze-shifts, but also provide critical input to reinforcement mechanisms in the basal ganglia which are instrumental in the acquisition of novel actions. These two important research themes will be brought together in the following study.

The experimental question relevant to the IM-CLEVER project would be: Can visual stimuli presented into the 'blind' field of unilaterally V1 lesioned monkeys also serve to reinforce the acquisition of novel actions? This question is directly analogous to that proposed for the IM-CLEVER experiments using masked stimuli, but with cortical lesions replacing the stimulus manipulation of masking. In other words, it addresses the same vital question of whether subcortical routes to the basal ganglia are sufficient support learning of novel actions independent of any input from cortical visual systems (as we suppose they do).

The study will be an eye-movement version of our current joy-stick task. Briefly, using eye-movements, the subject will have to discover an unseen location on the screen which, when the eyes move into it, triggers the presentation of a visual stimulus signifying the impending presentation of reward. (Another way of thinking about this would be as a visual version of the Morris water maze). The reward delivery would be delay for several seconds so the task has to be learnt from the secondary reinforcement properties of the visual stimulus rather the primary reinforcement properties of the reward delivery (which may be associated with auditory and/or somatosensory rather than visual stimuli). Each trial would begin with a stimulus identifying a fixation or start position the location of which will be randomly varied (this is needed so we can be sure he is learning the WHERE version of novel action acquisition). Thus, when the animal successfully saccades to the fixation stimulus it would disappear and the animal would then be free to move its eyes to 'search' for the unseen 'reinforcement area'. When the eyes move into this area, we will know that they have moved into a spatially defined 'reinforcement location' (the monkey are already fitted with corneal coils which measure the exact position of the eyes), which means we can present the visual reinforcer either into the normal or the blind field. The question then would be whether the monkey can learn WHERE the reinforcement location is (i.e. where he has to move his eyes to be reinforced) as effectively when the visual reinforcer is presented to the 'blind' or the visually intact field ?

There would be many important practical advantages of this study:
1) It takes full advantage of an existing population of V1 lesioned subjects, so no further animals would be required.
2) Minimal additional training would be required. The monkeys already know how to saccade to a fixation stimulus and they are used to 'searching' spontaneously in the visual field. The only difference in this case is the screen would be blank until the eyes moved into the critical, experimenter-defined reinforcement region, at which point the visual reinforcer would be presented.
3) All the advantages of the joy-stick task (see Redgrave et al and Stafford et al chapters in the IM-CLeVeR book delivery) would apply to the eye-movement version .i.e. it can be easy or difficult depending on the size of the reinforced region; when they have learnt one location (i.e. they move directly to the reinforced region irrespective of the location of the initial fixation stimulus), it can simply be moved to a new one; etc, etc.
4) All the software (MatLab) for this task is already developed in the USFD laboratory and would be readily available for this study
5) Depending on how rapid progress is, it would be possible to modify the task to become a WHAT version of action acquisition i.e. the animal would then have to discover WHAT movement from the fixation point is required to trigger reinforcement (which would be the same movement, irrespective of the varied location of the initial start/fixation position....i.e. opposite of the WHERE version where different movements are required to attain the 'reinforced location'.

Dependent measures of the study would of course would be the animals' behavioural performance, but also it will be very interesting to record the activity evoked by the visual reinforcer in the superior colliculus and/or in the dopaminergic neurones in the ventral midbrain. In this way the critical association between the sensory reinforcement used in the IM-CLEVER human studies of action acquisition and the activity of sensory-evoked changes in the activity of basal ganglia dopamine neurones when monkeys engage in a similar task can be made

Strand 1 represents a development of the USFD experiment programme according to development in theory (Redgrave etl al 2010). Strand 2 represents an strategically advantages collaboration with external collaborators which could directly address one of our existing experimental questions in a

novel way. This collaborative synergy would allow information exchange between projects and provide important biological insights which could be used by the IM-CLEVER project without involving any use of IM-CLEVER funds.

Introduction to the experiments with monkeys and children (UCBM & UCP)

When studying motivation in nonhuman animals distinguishing food-seeking exploration from 'generic' exploration is problematic. According to White (1959) acquisition of competence can explain the common biological significance of the exploratory behaviors through which the animal learns to interact effectively with its environment. Previous studies have shown that nonhuman primates learn to efficiently manipulate mechanical puzzle whose solution is not rewarded with food or water (Harlow 1950, Harlow et al., 1950). Exploration in chimpanzees has been studied by Welker (1956), who put several pairs of objects in front of them in order to observe the course of their interest towards them and to investigate whether the chimpanzees would show preferences. Size, brightness, heterogeneousness were important in eliciting interest towards the stimuli, and greater time was spent to interact with objects that could be moved, changed, or could emit sound and light than objects that could not. It has also been suggested that that actions could be reinforced by the opportunity to exert control over the environment (Glow et al., 1972; Glow and Winefield, 1978). Our experiments, inspired by the above studies, are designed to investigate the nature of intrinsically motivated cumulative learning in capuchin monkeys and human children. In particular we aim to assess whether affordances drive versatile learning.

The rationale underlying these experiments was that exploratory acts, not instrumental to achieve particular goal other than performance of the acts themselves, improve subjects' capacity in subsequent problem solving tasks which require the proficiency acquired during free exploration. Specifically, the experiments described below were aimed at evaluating whether tufted capuchin monkeys (*Cebus apella*) learn the affordances of novel objects (i.e. the different ways in which they can be manipulated and the unexpected consequences of particular manipulations) on the basis of their intrinsic motivation to explore them in the absence of obvious reward.

To this end, capuchin monkeys were tested with an experimental apparatus ("board") containing 3 buttons or 3 mechatronic modules (Taffoni et al., in prep.). The board was built in such a way that manipulation of buttons or modules produces a specific set of pre-programmed audio and visual stimuli along with the selective opening of one of the 3 boxes placed in front of the buttons/modules.

Experimental protocols used with monkeys and children

Introduction

Behavioural sciences is a term that encompasses all the disciplines which explore the activities and the interactions among organisms in the natural world. It involves the systematic rigorous analysis of human and animal behaviour through controlled experiments and naturalistic observations [1]. Behaviour is anything that a person or an animal does which can be observed and measured. There are several approaches to studying behaviour [2]. Especially in the past, while psychologists focused on the proximate causation of behaviour, and general processes of learning in a few animal species, namely those that better adapted to laboratory conditions, ethologists were typically interested in studying the ultimate causation of behaviour especially in nature where spontaneous behavior and the role played by environment could be better appreciated. Nowadays these two

fields are more integrated and neurosciences successfully contribute to clarify the neural correlates of behavior [3]. All the above disciplines require precise methods and tools for quantitative assessment of behaviors, possibly monitoring different levels of analysis, so to integrate them.

Materials and Methods

Experimental setup: description of the IM-CLeVeR Mechatronic Board

The mechatronic board is an innovative device specifically designed for inter-species comparative research on intrinsically motivated cumulative learning in children and non-human primates [1]. 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.
The mechatronic board is the result of a multidisciplinary design process, which has involved Bio-engineers, developmental neuroscientists, primatologists and roboticians to identify the main requirements and specifications of the platform.
The main requirements, which guided the design and fabrication of the board, are the following:

- to promote both intrinsically and extrinsically motivated actions that is, respectively, curiosity driven and rewarded actions. Moreover, it is essential for the purposes of the IM-CLeVeR project that the board allows performing actions capable of promoting a cumulative learning process;
- to embed non-intrusive technologies, which have to be ecological, small and light enough to fit the objects that will be manipulated by capuchin monkeys and children.
- to be equipped by instrumented interchangeable objects stimulating different kinds of manipulative behaviours and allowing to record several kinds of actions (e.g. rotations, pushing, pulling, repetitive hand movements, button pressing, etc);
- to record synchronized multimodal information for behavioural analysis and allowing to generate a set of different kinds of complex stimuli: visual, acoustic, and cognitive;
- to reward a reprogrammable set of actions, dispensing food for monkeys and small toys or stickers for children; the rewards are inserted in boxes which are made of transparent material, to allow the subjects to see when there is the reward inside them, but designed to prevent the subjects from opening the boxes when they do not perform the right action on the modules (rewarded action);
- to be made of materials, mechanisms, and electronic components robust enough for monkeys and children;
- to prevent any manipulation or interaction which would be potentially dangerous for the subjects or the board itself.

To easily reconfigure the experimental setup according to the requirements detailed above, a hierarchical, three level architecture was adopted (see Fig. 1):

1. The Physical Level, made by the physical interfaces subjects can directly interact with: push-buttons and mechatronic modules, and reward-releasing mechanisms. In particular, the mechatronic modules are composed of: (i) a mechanical interface with different affordances, eliciting different kinds of behaviours; (ii) a sensing core capable of measuring kinematic data during the interaction with the subjects; (iii) a communication interface (based on a I2C communication bus) that transfers the acquired data to the middleware level (hereafter called microcontroller-based level). On the other hand, the reward-releasing mechanisms are designed to allow catching the reward (food for monkeys and small toys or stickers for children) only when the correct action (i.e., the action that the experimenter has associated with the reward) is performed on the mechatronic modules; if other actions are performed, the reward remains visible but inaccessible.
2. The microcontroller-based middleware level. This level manages the low-level I/O

communication with the mechatronic modules and push buttons, the reward-releasing mechanisms, and audio-visual stimuli. The communication between the remote laptop and the microcontrollers is based on 4 different serial RS232 connections.

3. The high-level control and supervision system. A control software running on a remote laptop manages and supervises the acquisition and the arbitrary association between action and outcome

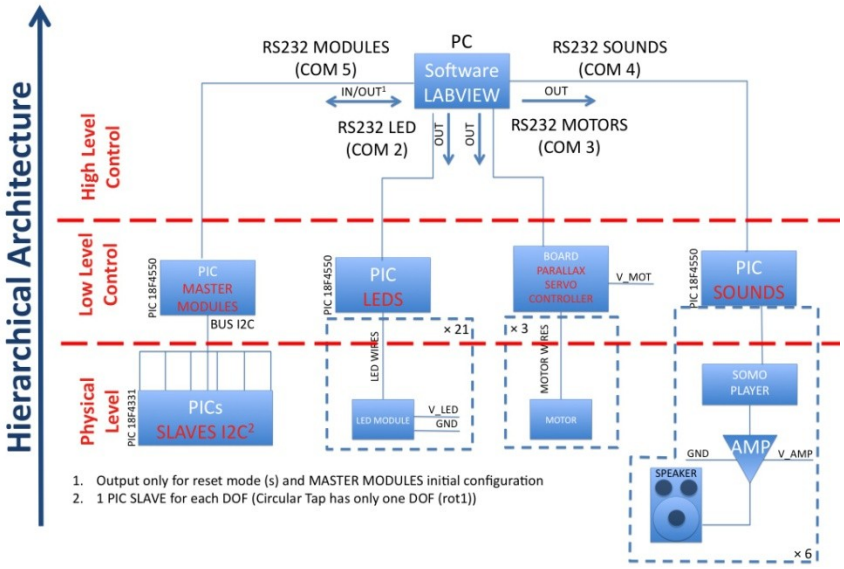


Fig. 1Hierarchical architecture of the board: physical level made by the interfaces with subject; local low-level control microcontroller-based; high level control running on a remote laptop

The mechatronic board has been designed and built in two versions for experiments on capuchin monkeys [4] and children. The two versions of the board are slightly different to take into account the differences between the two groups of subjects. The monkey version of the board is heavier, bigger and made by waterproof (since monkeys could urinate on it), non-varnished materials (since they like to remove the paint with their teeth or nails) which are chosen

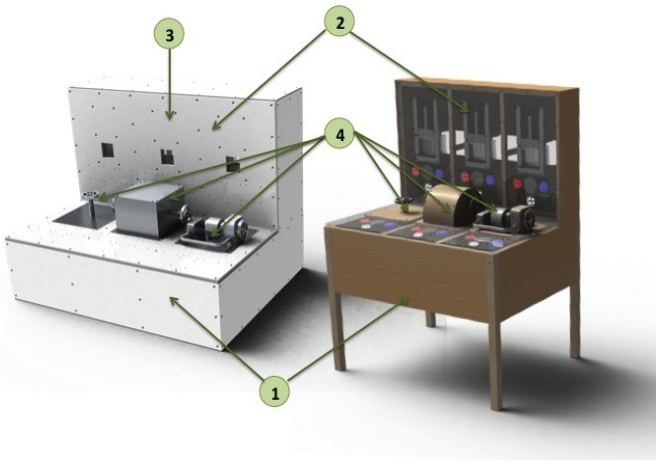


Fig. 2 The mechatronic board for monkeys (left) an children (right) : (1) planar base; (2) reward releasing unit; (3) local wide-angle camera (only in the monkey board); (4) mechatronic modules. The stimuli/reward system is not visible by the subjects, and it controls the aperture and closure of the reward boxes as well as the visual and acoustic stimuli.

to be robust enough to resist to typical monkey actions such as hitting, rubbing, biting. On the other hand, the baby version is similar, but scaled in dimension and made mainly of wood (see Fig.2).
Both the systems consist of the following components:

1. One planar base (overall dimensions are: 800x600x200 mm for monkeys board, 650x500x450 mm for children board): it is provided of three slots (200x200 mm; 180x180 mm) where the push-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 subjects can always see what is inside. The rear face of board is provided of proper openings, allowing the experimenters to easily insert the reward.
3. A local video-camera, embedded on top of the reward releasing unit in the monkey version monkeys, or an external camera for children, allows recording videos of the work space during the experiments.

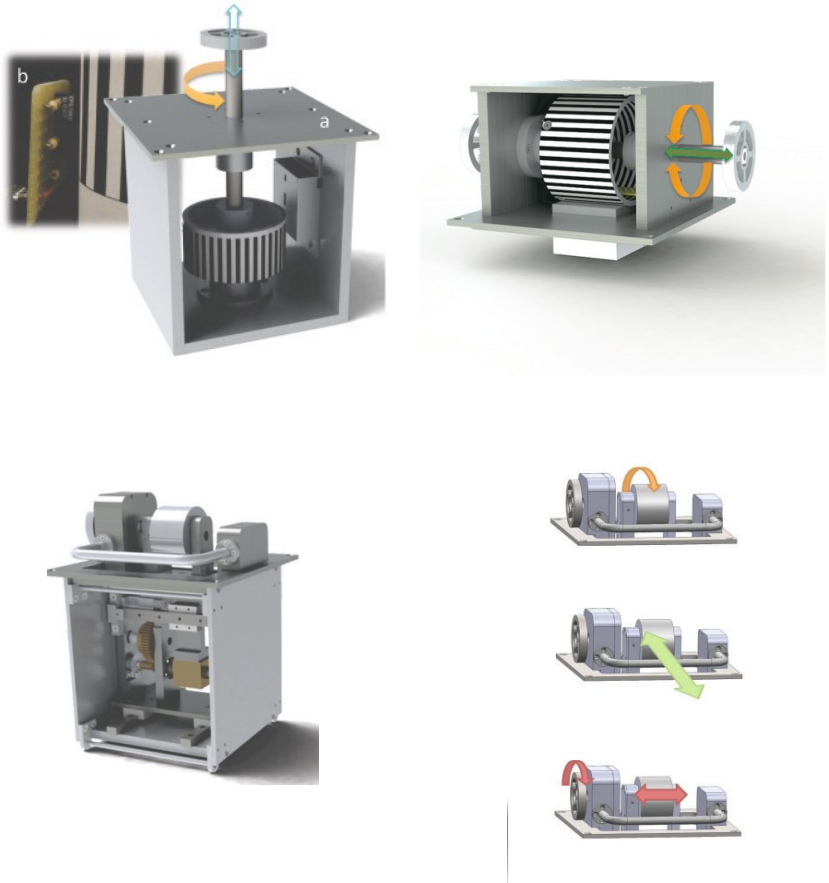


Fig. 3 (Up) On the left Circular Tap: (a) overall layout; (b) encoder electronics; On the right Fixed prism: the frontal wall has been removed allowing to see inner mechanism; (bottom) 3 Dof cylinder: overall layout (left); degrees of freedom (rigth): central wheel rotation; horizontal translation (middle); translation of the central wheel driven by lateral circular handle bottom

4. Push-buttons and mechatronic modules: each of them is provided by a specific set of sensors and a local microcontroller unit, which 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 makes reconfiguration of the system possible, allowing to easily collect data from the different peripherals. For the mechatronic modules, only optical sensors were used in order to physically separate electronics to physical interfaces, in order to avoid any direct interaction between subjects (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 in 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 push-button modules and three complex mechatronic modules. The push-buttons modules allow detecting simple pressing action or they can be programmed to respond to more complex interaction, such as multiple consecutive presses or a hold press (the time interval can be arbitrarily set by the high level control system). 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 subject rotates the central cylinder or translates it using the horizontal handle, or mediated by a inner mechanism, which translates the rotation of the lateral wheel in an horizontal translation of the cylinder along its main axis. Figure 3 shows the affordances and the degrees of freedom of the three mechatronic modules.

5. Stimuli and reward system: the whole platform is equipped with a set of different stimuli (acoustic and visual) to provide various sensory feedbacks associated to the manipulation of 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 used during the experiments were chosen from a large database of natural and artificial sounds. The visual stimuli consist of a set of 21 independent multicoloured 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 subject can retrieve the reward only when he/she performs the correct action on the mechatronic modules. The reward releasing mechanism was designed to be not backdriveable (so that the subject cannot force the opening) (see fig. 4). A Parallax Continuous Rotation Servo motor (maximal torque: 0.33 Nm) has been used to drive the 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 subjects hand is caught in the sliding door; furthermore, since the mechanism is not backdriveable, it does not allow the subject to force the opening of the sliding door. The action-outcome association is managed by the high-level control system and is fully programmable according to the experiments requirements.

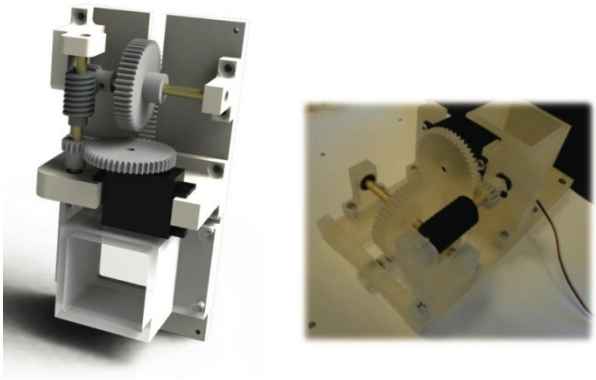


Fig. 4 Reward/releasing mechanism: on the left rendering of the mechanism; on the right, the developed mechanism

All the electronics of the microcontroller-based middleware level has been integrated in a single motherboard, which could be easily embedded into the planar base, and connected to the Audio/video stimuli boards and to the mechatronic modules using 10-way flat cables (see fig. 5).
In figure 6 the pictures of the two version of the board for monkeys and children are shown.

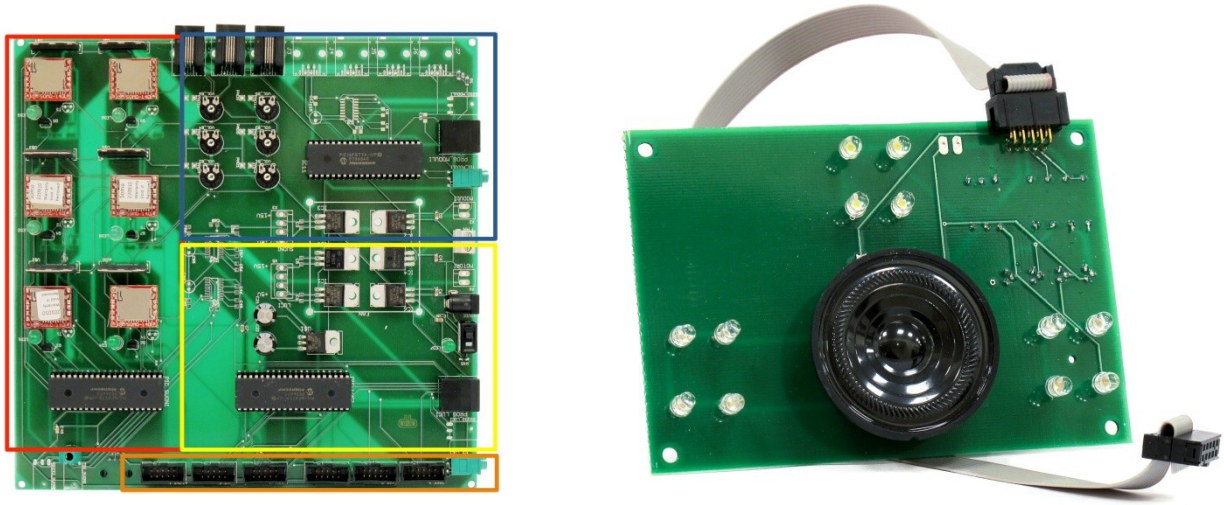


Fig. 5: (Left) Motherboard with the electronics of the microcontroller-based middleware system: PIC master for mechatronic modules (blue); PIC for Led control (yellow); PIC for sounds control (red); connectors for audio/video stimuli board (orange). (Right) audio-visual stimuli board

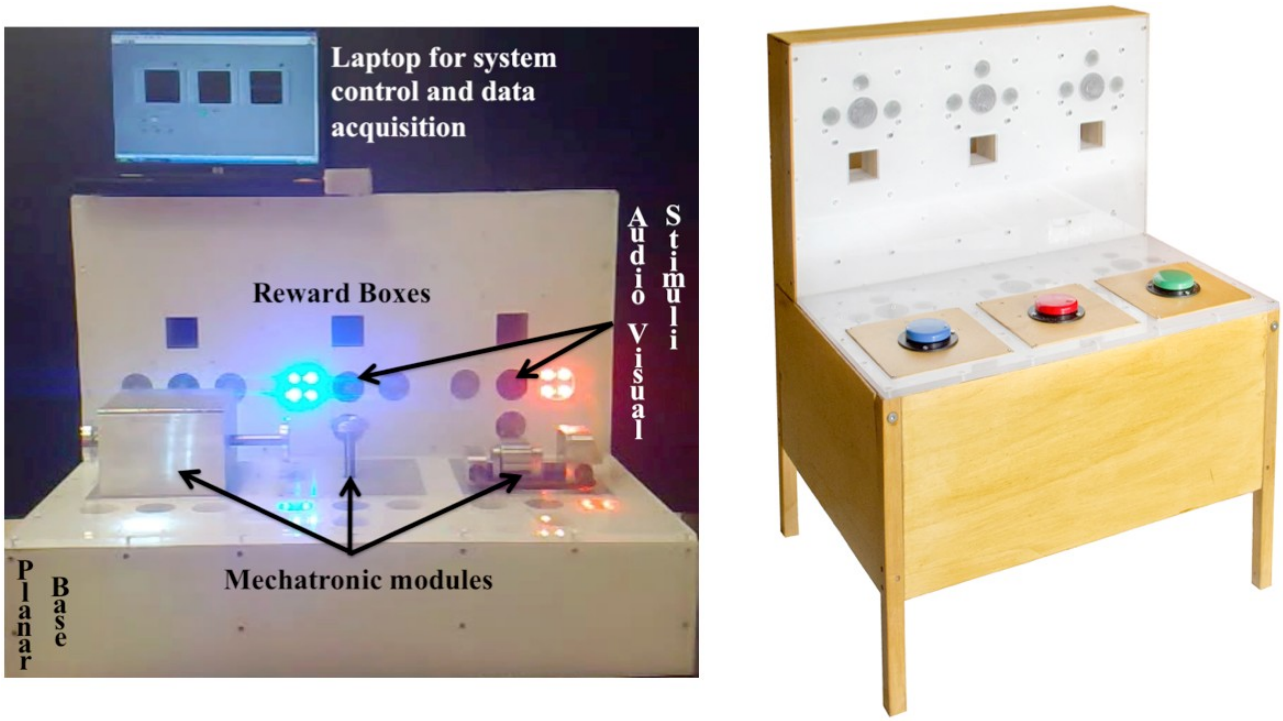


Fig. 6: The mechatronic board in the two different versions for monkeys (left) and children (right).

Children Experimental protocol with the push buttons

In order to test the board and to optimize the experimental protocol (in particular the verbal instructions given to the children) a pilot study on two children (two female children of 24 and 46 months of age, respectively) has been performed at the day-care centre ‘La Primavera del Campus’,

(Università Campus Bio-Medico, Rome, Italy), using the mechatronic board equipped with three push-buttons (Blue, Green, and Red). Children have been recruited by explaining to the parents the general goals of the project and asking them to sign an informed consent (approved by the IRB of Università Campus Bio-Medico).
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 say anything about what the board actually does.
The experimental protocol is divided in two phases: a training phase and a test phase. The basic goal of the protocol is to assess whether a child can use a motor skill that he/she has acquired during the training phase (push a button in a way that opens a box) to retrieve a reward in the test phase. During the training phase the child discovers "by chance" that he/she can open the boxes. In the training phase the child can freely explore the board and its functionalities. The board is programmed to react to each single press of the buttons with both visual and audio stimuli, and to open the reward boxes when a button is hold pressed for more than two seconds (rewarded action). The single press makes the lights close to the button to turn on and causes a single xylophone note to sound (three different notes are set for the three buttons). On the other hand the rewarded action produces the opening of one box (which is always empty in the Learning Phase), the lighting of the box lights and the light inside the box, and at the same time generates a sound of an animal cry (one for each button: a rooster, a frog and a cat). For all subjects, the Blue Button (BB) opened the Left box (Lb), the Red Button (RB) the Rigth box (Rb) and the Green Button (GB) the Central box (Cb) (see fig 7).
The Learning phase lasts about 10 minutes and is followed by the second phase (hereafter called Test Phase).
In the Test Phase the reward (a sticker) is shown to the child and then randomly placed in one of the three closed boxes, where it is clearly visible to the subject. The child is only asked to retrieve the sticker, without adding any other suggestion on what action is associated to box opening. As in the Training Phase, the reward can be reached by pushing and holding the associated button for more than 2 seconds. The other stimuli are set as in phase 1.
Once the subject opens the box and reaches the reward, it is given to the child as a prize for his/her success. The Test Phase ends after 9 successful openings (three for each box) or after 40 min.
The subjects are divided in two groups: the Experimental Group and the Control Group. The protocols for the two groups differ only in the Training Phase: while in the Experimental Group the rewarded action causes the opening of the associated box also in the training phase phase, in the Control group the boxes do not open in the training phase. All the other audio-visual stimuli are set in the same way in both groups.

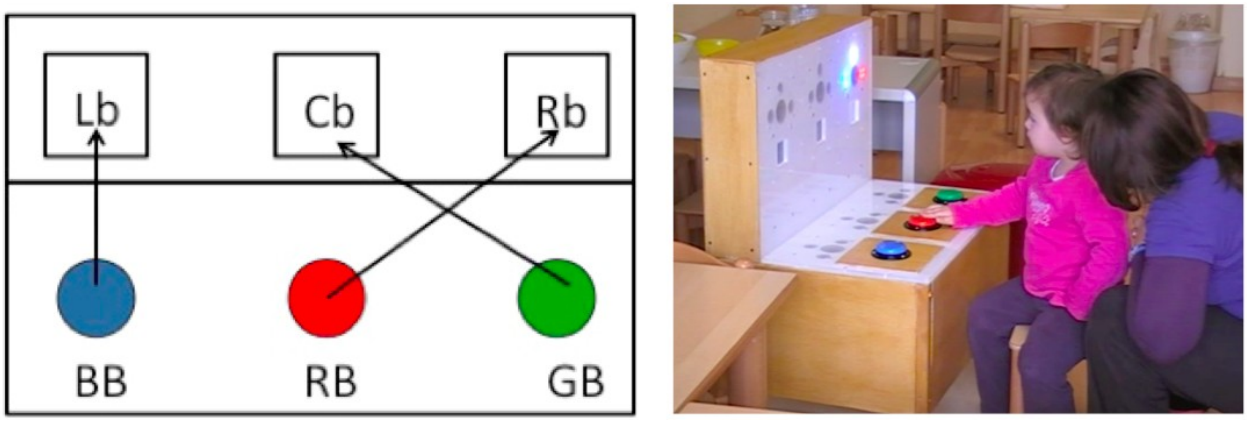


Fig. 7. Schematic representation of the arrangement of buttons and their association with boxes from the perspective of the user (left). A snapshot of the experiment during the Training Phase (right).

Children Experimental protocol with the mechatronic modules

The experimental protocol for the experiments with the mechatronic modules has the same structure and organization (two phases and two subjects' groups, same spatial association between modules and boxes as shown in fig. 7 - left) as the one previously described for the push-buttons. The only differences are related to the audio-visual stimuli associated to the more complex actions that can be performed with the modules.

In particular the stimuli/reward outcome are programmed for the three different modules as follows:

1. Circular tap: tap rotation makes the lights close to the module to rotate in the same way; lift up the handle generates a single xylophone note; rotating the tap for at least 360 degrees in 5 seconds opens the corresponding box, switches on the box lights, and plays the related animal cry (rewarded action).

2. Fixed prism: the rotation of the module shaft makes the lights close to the module to rotate in the same way; rotating the shaft for at least 360 degrees in 5 second generates a single xylophone note; pulling or pushing the shaft for more than 1 half of its total stroke in 5 seconds is the rewarded action for this module (lights, box opening, and animal cry).

3. 3 dof Cylinder: the rotation of central cylinder makes the lights close to the module to rotate in the same way; the translation back and forth of the modules generates the xylophone note; the rotate the lateral handle for more than 1 half of its total stroke in 5 seconds is the rewarded action for this module (lights, box opening, and animal cry).

Experiments with capuchin monkeys (CNR-ISTC-UCP)

1. Pilot study: Experiment with the button board

The pilot study specifically aimed at testing the functioning of the board with capuchins monkeys, a species well known to be manipulative and destructive when dealing with objects and food items (Fragaszy et al., 2004).

The data presented here refer to the button condition that preceded the use of the mechatronic objects and whose action-outcome associations were assumed to be less demanding for monkeys to learn, than mechatronic modules which present more affordances.

During the pilot study systematic data were collected on the monkeys' initial response to the mechatronic platform and on the monkeys' manipulation of the buttons. These information were important to understand whether the apparatus and/or the procedure needed any changes before to start the planned experiments on implicit learning.

Subjects and experimental apparatus

The subjects of the pilot experiments were 3 adult capuchin monkeys (Pedro, Robiola and Robin Hood) hosted at the Primate Centre of the Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy.

Capuchins were tested individually in an indoor enclosure (5 m^2 x 2.5 m high). Each subject was separated from the group solely for the purpose of testing, just before her/his testing session. Subjects were

not food deprived and water was freely available at all times. The board had 3 buttons of different colors (white, black, and red), placed at about 25 cm apart from one another along the same line (see Fig. 1), that could be discriminate by trichromatic and dichromatic subjects (capuchin monkeys males are all dichromats, whereas females could 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 3 boxes.

Experimental design and procedure

The pilot experiment included two phases. In Phase 1 the correct action performed by the subject (i.e. pushing a button at least once) produced a specific combination of audio and visual effects together with the opening of one box. The box did not contain any reward. Phase 1 lasted for 20 min. In Phase 2, the reward (one peanut) was located in one of the three boxes in clear view of the subject. The reward could be obtained by pushing the associated button (see Fig. 1). Each subject received 9 trials and the reward position was balanced across boxes. Phase 2 ended after 9 trials or when 40 min elapsed, whichever came first.

For all subjects, the white button (WB) opened the central box (CB), the black button (BB) the left box (LB) and the red button (RB) the right box (RB) (see Fig. 1). Thus, the spatial relation between button and associated box was crossed for WB and BB and frontal for RB. The experimental board was placed on the floor of the experimental cage, attached to the front wire-mesh of the cage. The pilot experiment was videotaped by a camera (Sony Handycam, DCR-SR35) and local wide-angle camera embedded in the board. The ELAN software allowed to synchronize the videos obtained by the two cameras.

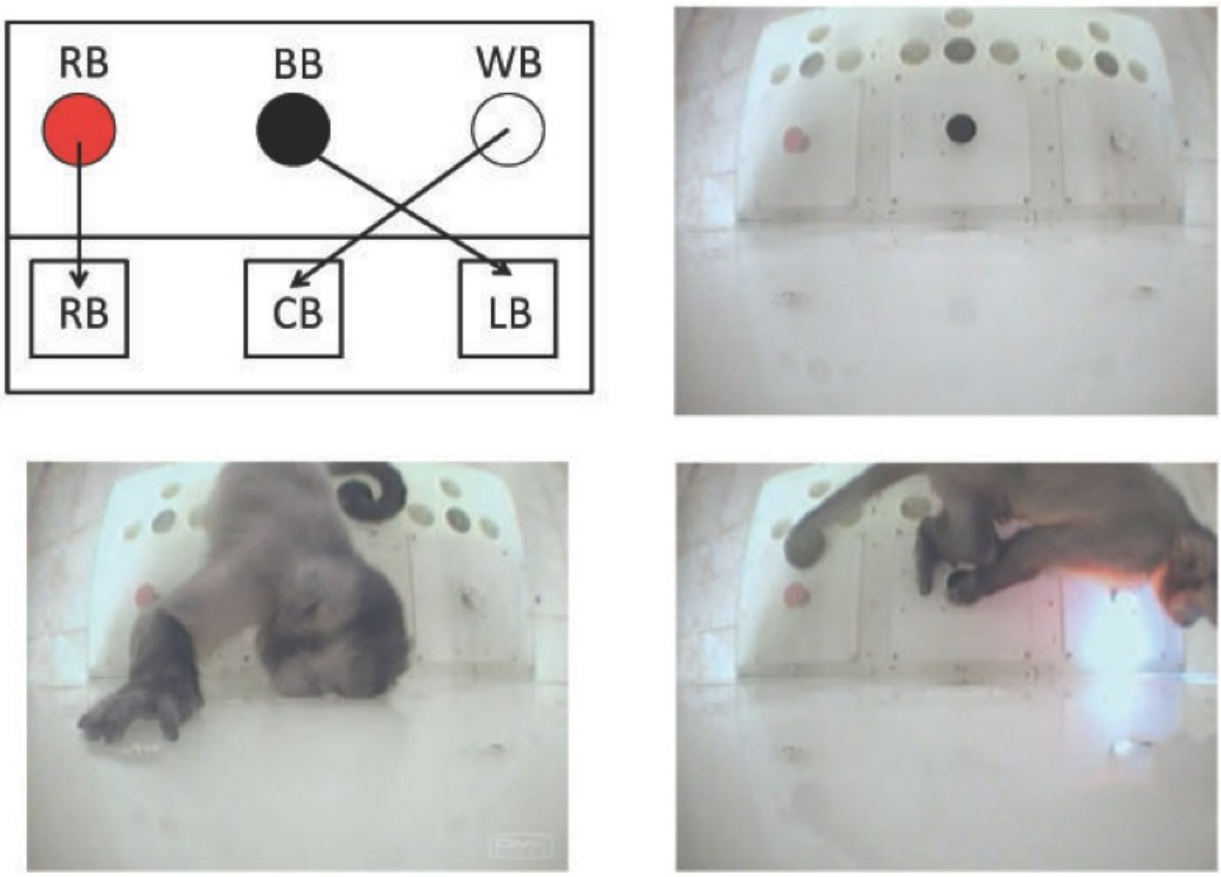


Fig. 1

(Top left) Drawing of the disposition of buttons and their association with boxes. (Top right) Screenshot taken by the board embedded webcam showing a view of the workspace. (Bottom left) A monkey visually explores the central box. (Bottom right) A monkey has pushed the black button opening the left (lighted) box. Note that the subject is watching this box.

Results

Phase 1

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. Her average time during which she held the button pressed was 0.17 sec (SE: ± 0.008). The overall mean time in contact with the board was 5 min and 5 sec and the value varied across subjects (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. Boxes close distance exploration (within 10 cm) never occurred for Robiola, whereas Pedro did it once and Robin Hood 8 times. The overall mean scratching rate, (used as a behavioral measure of stress) occurred at 0.4 events/min (SE: ± 0.02).

Phase 2

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 ± SE: 19 min and 10 sec ± 2.76) and manipulated each button much longer (Mean ± SE: 40 sec ± 8.03). Scratching occurred at a higher rate than in Phase 1 (Mean± SE: 0.6 events/min ± 0.05).

Table 1 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 pushed 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.

Table 1. Measures collected by the board during Phase 2 for each box-button association. The number of pushes and the holding time refer to button, whereas the number of incorrect responses refers to the rewarded box of the corresponding column.

	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 ±SE	1. 9 ±0.8	0. 8 ±0.3	1 ±0.25
Mean number of incorrect responses per subject per trial ±SE	1.2 ±0.2	3. 7 ±0.7	1. 2 ±0.3
Mean holding time per subject per trial ±SE	0.2 ±0.05	0.25 ±0.03	0.3 ±0.11

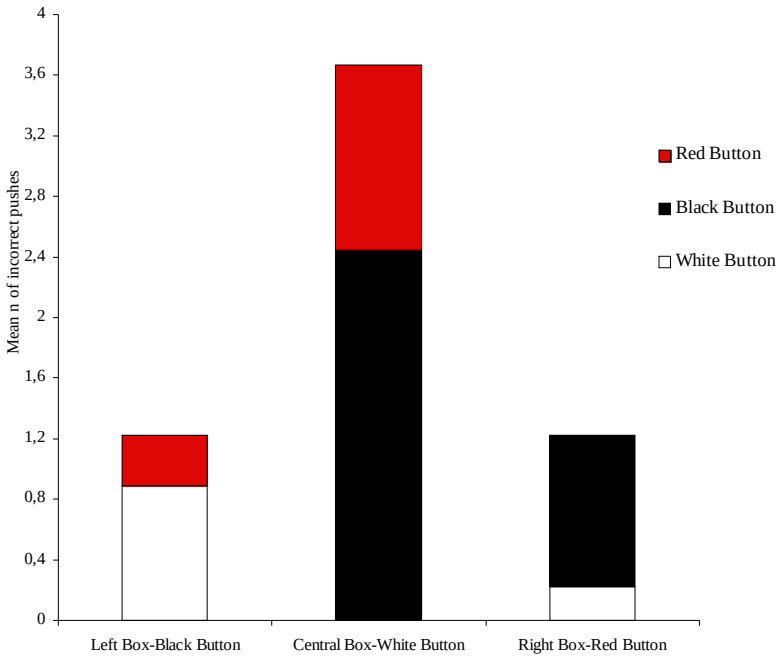


Fig. 2.

Mean number of incorrect pushes (per subject per trial) performed while the reward was in the left box, in the central box, and in the right box (x axis).

Discussion

So far, the results of the pilot study suggest that capuchin monkeys were not particularly interested by the buttons *per se*. In fact, in Phase 1 they spent relatively little time in contact with the board; moreover, the pushing action occurred rarely and apparently not on purpose. However, the presence of a reward during Phase 2 increases monkeys' interest toward the board and triggered a variety of behaviours, such as visual exploration, time in contact with the board and pushing the buttons. These behaviours may eventually lead capuchins to learn specific action-outcome associations.

The association between boxes and buttons in the crossed condition was perceived 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 subject position 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, our 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 afforded only the action of pushing. On the basis of these results, we proceeded to test the experimental subjects only with the board containing mechatronic modules.

Experiment with the mechatronic board

The experiment involves two phases. In Phase 1 (exploration phase) the subjects are free to interact with the modules without receiving any kind of reinforcement whereas in Phase 2 (problem solving task or testing phase) the same modules become devices to obtain the reward.

The three mechatronic modules (Circular Tap, Fixed Prism and 3dof Cylinder), all designed in order to elicit interest and exploration, differ in terms of the number of actions that they can afford. In both phases when the subject explores a given affordance of a module, the corresponding box opens for and produces

interesting outcomes, i.e., light and sound which differ among boxes.

Subjects

The subjects were 16 adult tufted capuchin monkeys (*Cebus apella*). Capuchins were tested individually in the indoor cage to which they have access through a sliding door from the adjacent outdoor cage. Each subject was separated from the group solely for the purpose of testing, just before each daily testing session. Testing was done in the morning/early afternoon. Subjects were never food deprived and received their main meal (monkey chow, fresh fruit and vegetables) in the afternoon. Water was freely available at all times. Half of the subjects received Phase 1 and 2 (Experimental group), the other half received a Familiarization Phase followed by Phase 2 (Control group). All the experiments were was videotaped by a camera (Sony Handycam, DCR-SR35) placed at 45-60 degrees from the front of the board and by the local wide-angle camera embedded in the board.

Exploration attitude assessment. Before starting with the experiment, 24 capuchins were tested with a single-affordance object, a door handle (Fig. 3) in order to determine their likelihood to explore. To this end, we measured the number of times subjects manipulated the handle and the overall time spent in contact with the board. The exploration phase lasted for 5 minutes and was video-recorded. On this basis, we categorized them in “explorative” and “non explorative” individuals and identified 12 “explorative” individuals. “Explorative” individuals were split between the “Experimental" group and the “Control” group. Other 4 individuals was randomly chosen from the “non explorative” group and assigned to each group. So, at the end, we had 8 subjects per group.

Fig.3

Screenshot of the door handle on the board (Left) and a capuchin monkey manipulating it (Right).

Methods

Phase 1

In this phase the interaction with the three mechatronic modules produced a pre-programmed set of audio and visual stimuli. When the subject explored a given affordance of each module (pre-determined by the experimenter, see Table 2), an empty box opened accompanied by a specific set of lights and sound.

A subject should open each box at least 5 times before proceeding to Phase 2. This criterion could be achieved in one session, or in a cumulative way during more training sessions. However, if a subject showed little interest towards the board (less than 460 sec of contact with the board; this measure corresponds to the mean time during which capuchins interacted with the board in the pilot experiment), he/she entered directly Phase 2. This was done to prevent habituation to the board (and further decrease the subject’s interest in Phase 2).

In order to avoid a novelty effect, the Control group familiarized with the board and with the empty

boxes (“Familiarization Phase”). This differed from Phase 1 only because boxes could not be opened.

Phase 2

Both Experimental and Control groups underwent Phase 2. In this phase, a reward (one unshelled peanut) was located in clear view of the subject in one of the three boxes. The reward position was balanced across the three boxes and presented in a pseudo-random order (the reward was not be placed on the same box more than two consecutive times). In each trial, a single box was rewarded and the reward could be obtained only by performing the correct action on the correct module, as programmed by the experimenter.

A trial ended when the subject performed the action that opened the box and retrieved the reward, or when 5 min have elapsed. In the latter case, the experimenter sent the subject in an adjacent cage, retrieved the reward from the box, and baited another box while attracting the subject’s attention. When the subject solved the task, the experimenter sent the subject in an adjacent cage, baited again one box and, as soon as the subject ended eating, attracted its attention to the baited box. A total of 9 trials were performed. There were two experimental conditions balanced both for type and number of affordances and spatial correspondence between module and baited box (frontal or crossed; for details, see Fig 4).

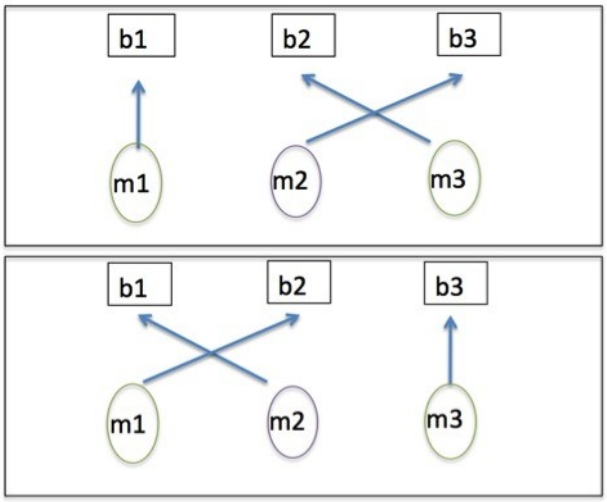


Fig.4

Experimental conditions. In Condition 1 (above) the association between box (b) and manipulandum (m) is crossed for the central double affordance manipulandum (m2) and for the three-affordance manipulandum (m3), while it is frontal for the double-affordance manipulandum (m1). In Condition 2 (below) the association between box and module is crossed for the two-affordance manipulanda (m1 and m2), while it is frontal for the three-affordance manipulandum (m3). The two conditions allowed controlling for the spatial relation between module and box (frontal or crossed) and the number of affordances (two vs. three) of the manipulandum. Moreover, the actions rewarded for m2 differ between conditions.

Table 2.

Type of actions rewarded in each of the two experimental conditions.

Condition 1	
module 1 (Circular Tap)	rotation of at least 360° in both directions (clockwise or anticlockwise)
module 2 (Fixed Prism)	pulling
module 3 (3dofCilinder)	rotation of the handle of at least 90° in both directions
Condition 2	

module 1 (Circular Tap)	rotation of at least 360° in both directions (clockwise or anticlockwise)
module 2 (Fixed Prism)	rotation of at least 360° in both directions (clockwise or anticlockwise)
module 3 (3dofCilinder)	rotation of the handle of at least 90° in both directions

Data scoring

The board scores the following data: latency to first exploration of each module and latency to first exploration of each affordance; task persistence (i.e., the time each participant manipulates the module); richness of investigation, i.e., number of different actions performed on the module as well as the number of times an effect (e.g., sound, light) is produced; frequency/speed of each performed action. Additional data were collected by scoring the following behaviours from the video-recordings of the two cameras: body contact with the board, box visual exploration (the number of times in which subject visually inspect each box from a distance between 0-10 cm from it) and frequency of scratching, a behavioural measures of stress.

These data allow us to assess whether (a) subjects exploring more during Phase 1 will perform better in the Phase 2 (b) objects presenting more affordances are more explored; (c) the experimental group will perform better than naïve subjects of the Control group who have never experienced the association between object manipulation and box opening (d) patterns of exploration (# of affordances exploited, order of exploration, preference, etc.).

Data analysis

Data were analyzed by using parametric statistics as data had a normal distribution. As we did not expect conditions to influence the behaviour of individuals in terms of latency to approach the board, arousal, and time in contact with the board, we analyzed their response only as experimental and control group by using t-tests. In order to reveal what factors affected individuals' performance during Phase 2, we applied an Anova test for repeated measures using the number of rewards obtained in each box as dependent variable and group, condition an their interactions as independent variables.

The distribution of errors performed during Phase 2 was analyzed by using Linear Mixed Models. LMM are a useful tool when dealing with data showing interindependency, as it may be the case of multiple datapoints per subject. Moreover, LMM deal successfully with unbalanced datasets, as several individuals could have undergone to multiple training and familiarization sessions in Phase 1. LMM control for differences in the individual contribution to the data set, allow unbalanced data to be analysed, control for the effect of independent variables on one another, and finally control for individuals identities using random factors. In this experiment, the number of errors performed by each subject during each session and for each box was entered as data-point (dependent variable), while the identity of boxes (left, center, right), group (experimental or control), condition (1 or 2), type of module-box association (frontal or crossed) and number of time each rewarded box was opened/illuminated during previous training/familiarization sessions were included as independent variables. The identity of subjects was included as random factor so as to control for between-subject variation.

Results with capuchin monkeys (CNR-ISTC-UCP)

Phase 1. Training and Familiarization sessions

General behaviours

During Phase 1, the time individuals spent in contact with the board did not differ among experimental and control groups (t_{14}=-1.55, P=0.24, Fig. 8). Similarly, no difference was found in the latency to the first approach to the board (t_{14}= 0.63, P=0.52, Fig. 9). No difference in the amount of arousal was found between groups, as revealed by the rate of scratching during the sessions (Mean ± SE Control group 0.25 ± 0.09 ev/min; Experimental group 0.11 ± 0.05 ev/min; t_{14}= 1.08, P=0.28)

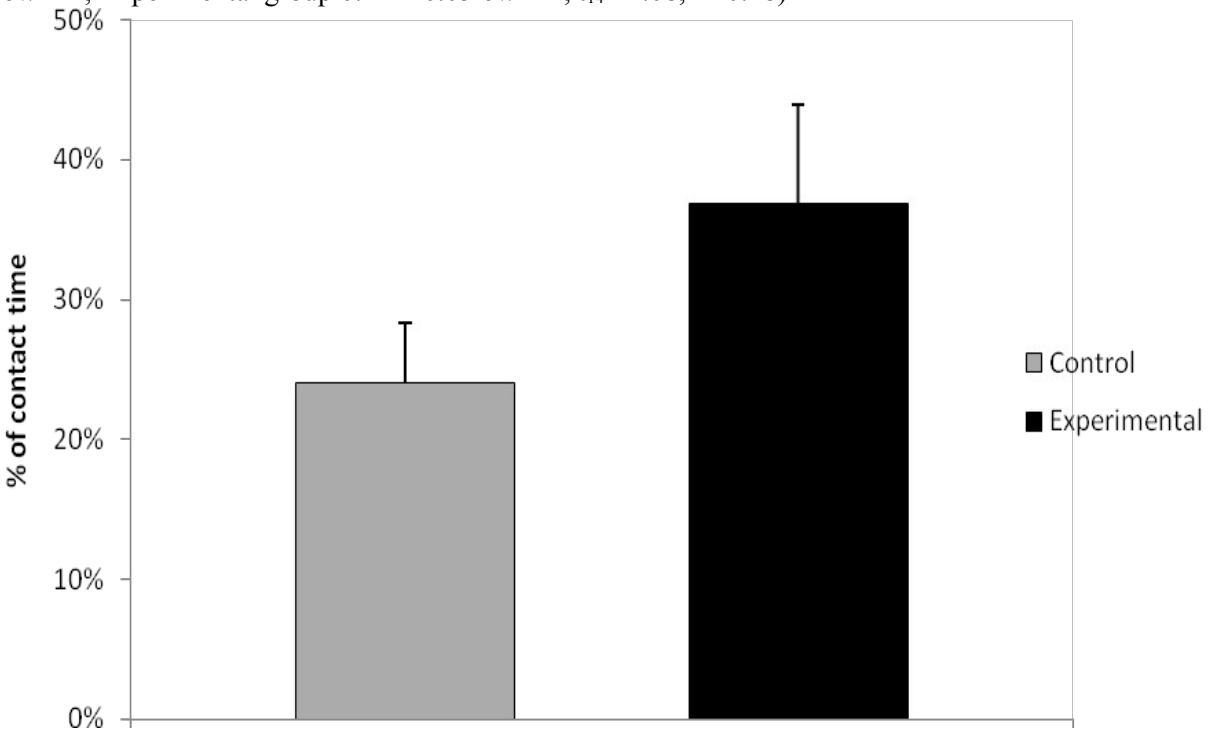


Fig. 8 Percentage (± SE) of time in contact with the board for the experimental and control group.

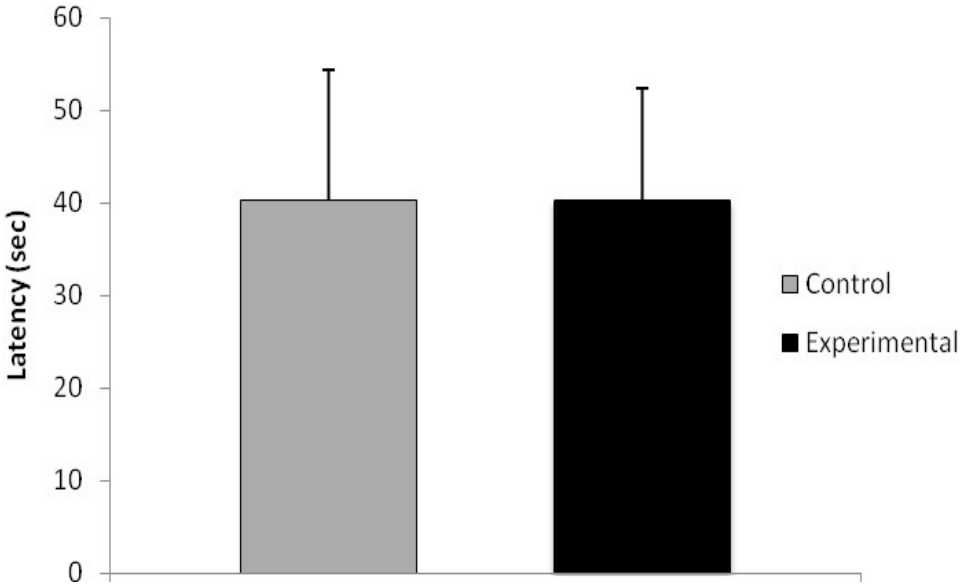


Fig. 9 Mean (± SE) latency to the first approach with the board for the experimental and control group.

Manipulation of the modules

The experimental group tended to manipulate the modules more than the control group (Mean ± SE: 46.8±17.4 sec and 15.6± 4.2 sec, respectively), but the difference was not statistically significant (Anova main effect group: F_2=2.931, P=0.109, Fig. 10). The two groups manipulated each module similarly (Anova group*modules interaction effect: F_2=1.078, P=0.354), and the overall manipulation time differed significantly across modules (Anova main effect of modules: F_1=4.544, P=0.019). Post-hoc tests revealed that the circular tap was manipulated significantly less than the 3dof Cylinder (post-hoc test: P=0.0092), while no difference was found for the other comparisons (Circular tap vs Fixed prism: P=0.254 ; Fixed Prism vs 3DofCylinder: P=0.809).

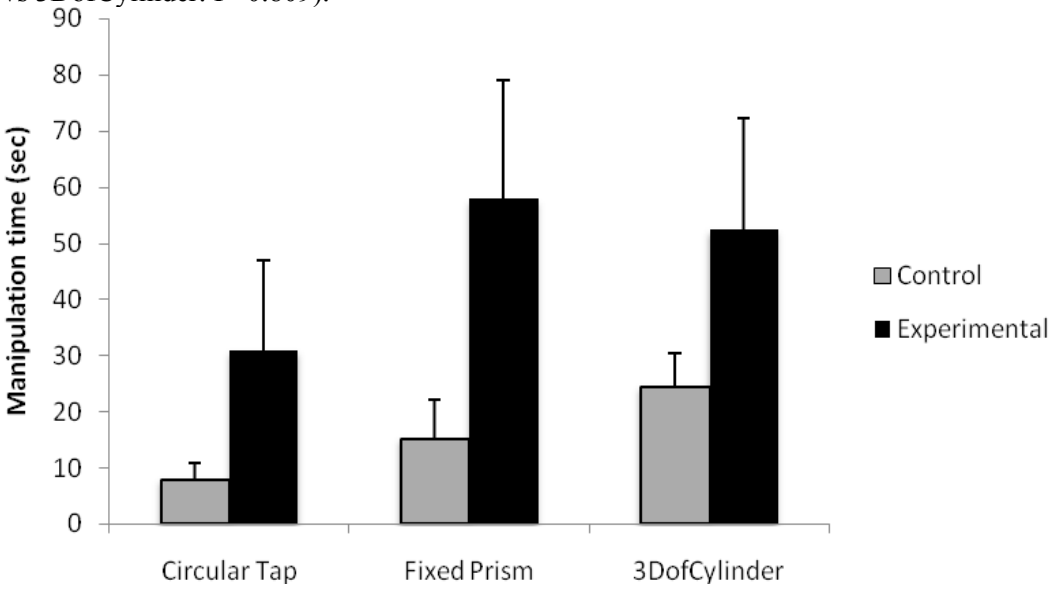


Fig. 10 Mean (± SE) manipulation time for each module for the control and experimental group.

Exploration of boxes

The experimental group did not explore the board boxes (Mean ± SE: 0.36± 0.16 ev/min) significantly more than the control group (Mean ± SE: 0.19 ±0.04 ev/min; t-test: t_{14}=-1.0278, P=0.3261).

Modules actions

Results showed that there was no difference in the frequency with which actions were performed by the control and experimental group for all of the modules (Anova, interaction effect of group*actions Circular Tap: F_1=0.579, P=0.460; Fixed Prism: F_1=0.055 P=0.818, 3DofCylinder: F_1=2.259, P=0.125, Fig. 11). However, there was a significant main effect of actions for the 3DofCylinder (Anova, main effect of actions: F_1=7.432, P=0.003). The push-pull action of the 3dofCylinder overall was performed significantly less than the corresponding wheel (post-hoc test: P=0.0164) and leverage action (post-hoc test: P=0.0085)

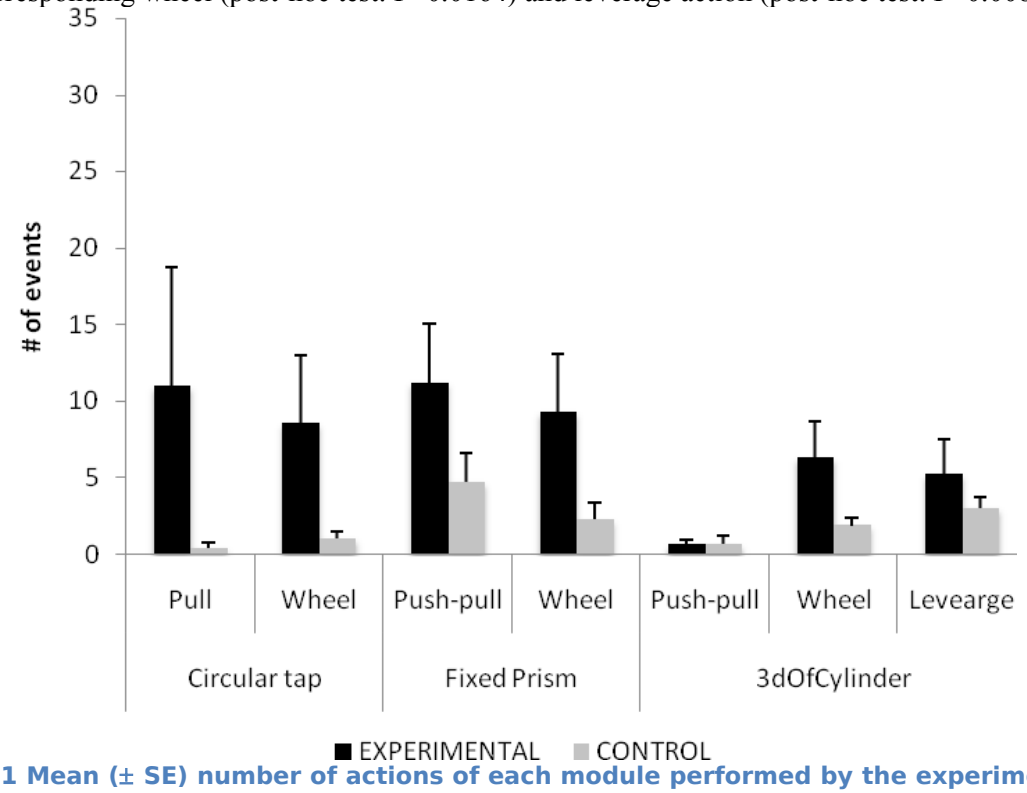


Fig. 11 Mean (± SE) number of actions of each module performed by the experimental and the control group.

Phase 2. Test

Groups' performance

Individuals of the experimental group did not obtain the reward more frequently than those of the control group (t-test: t_{14}=-0.554, P=0.589, Mean ± SE Experimental group: 0.66±0.15, Control group: 0.55±0.13). However, the experimental group had a percentage of success (in terms of number of times the reward was obtained) that positively correlated with the time spent manipulating the modules during phase 1. This correlation almost reached statistical significance ($t_{1,6}$=2.30, P=0.061). In contrast, no correlation was found for the control group (t_{14}=1.07, P=0.327, Fig. 12).

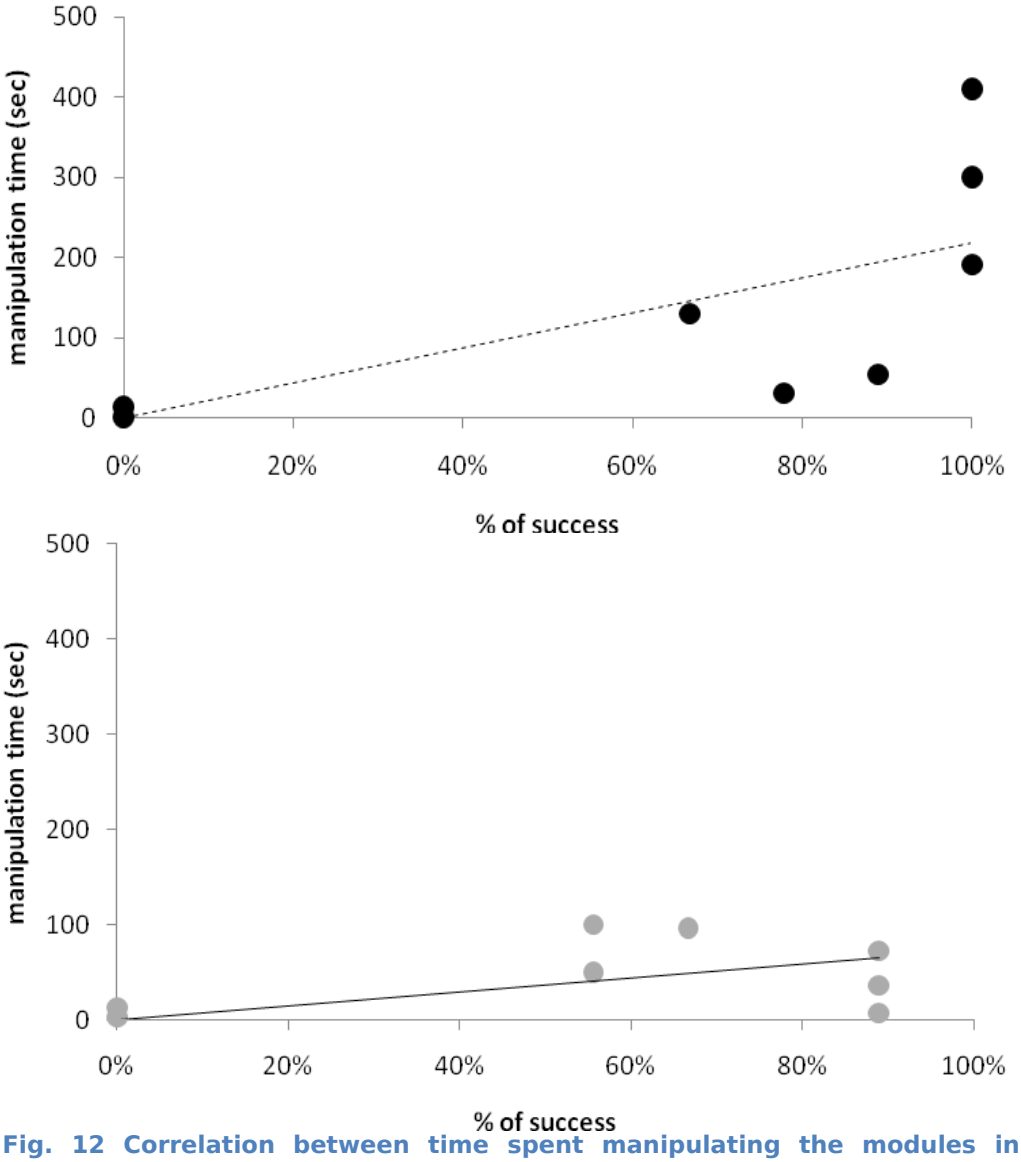


Fig. 12 Correlation between time spent manipulating the modules in Phase 1 and the percentage of success during Phase 2 for the experimental (above, black dots) and the control (below, grey dots) group. Data points represent individuals of the two groups.

Arousal

For both groups, arousal did not differ between Phase 1 and Phase 2, (Control group: t_{15}=-1.09, P=0.293; Experimental group t_{15}=0.31 P=0.754, Fig. 13). No difference in the amount of arousal was found between groups during Phase 2 (t test: t_{14}=-0.0196, P=0.9846)

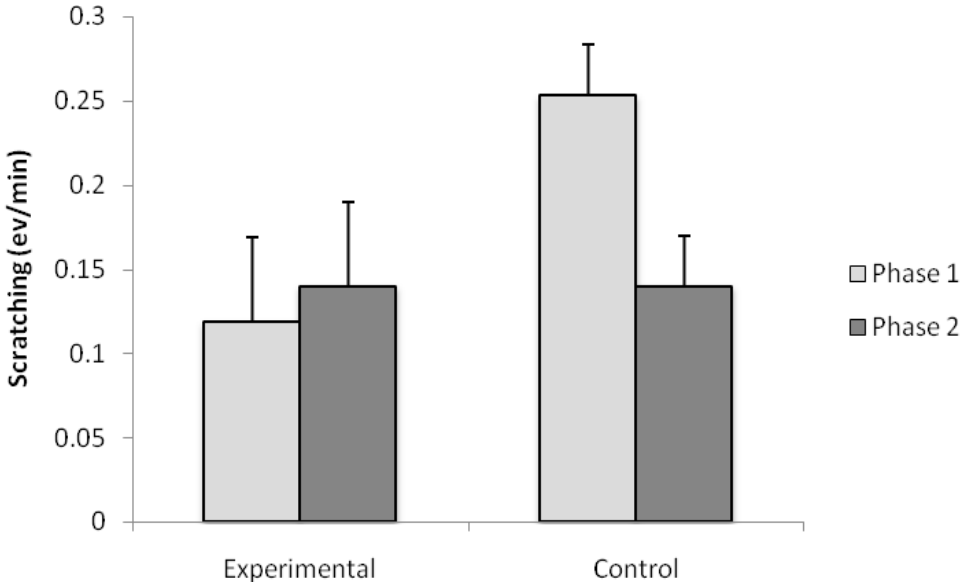


Fig. 13 Mean (± SE) rate of scratching for the experimental and control group during phase 1 and 2.

Influence on individuals' performance: success

Given that no difference was found between experimental and control groups, we combined the two groups for subsequent analyses. The number of successful trials did not differ between conditions (Mean ± SE: condition 1: 4.125 ±1.27, condition 2: 6.87± 1.07; t-test t_{14}=-1.55, P=0.164).

Success in frontal and crossed associations

The number of successful trials scored in the frontal box-module associations (Mean ± SE: 0.62 ± 0.11) did not significantly differ from that scored in the crossed box-module associations (Mean ± SE 0.60 ± 0.10; t-test: t_{15}=-0.286, P=0.778).

Success in crossed associations

The crossed associations between box and module could involve a) the central module or b) one the two lateral modules. The number of successful trials was significantly higher for the central module-box association than for the lateral modules-box association (central: 0.73 ± 0.11, lateral: 0.48 ± 0.1, t-test: t_{15}=2.535, P=0.02, Fig. 14).

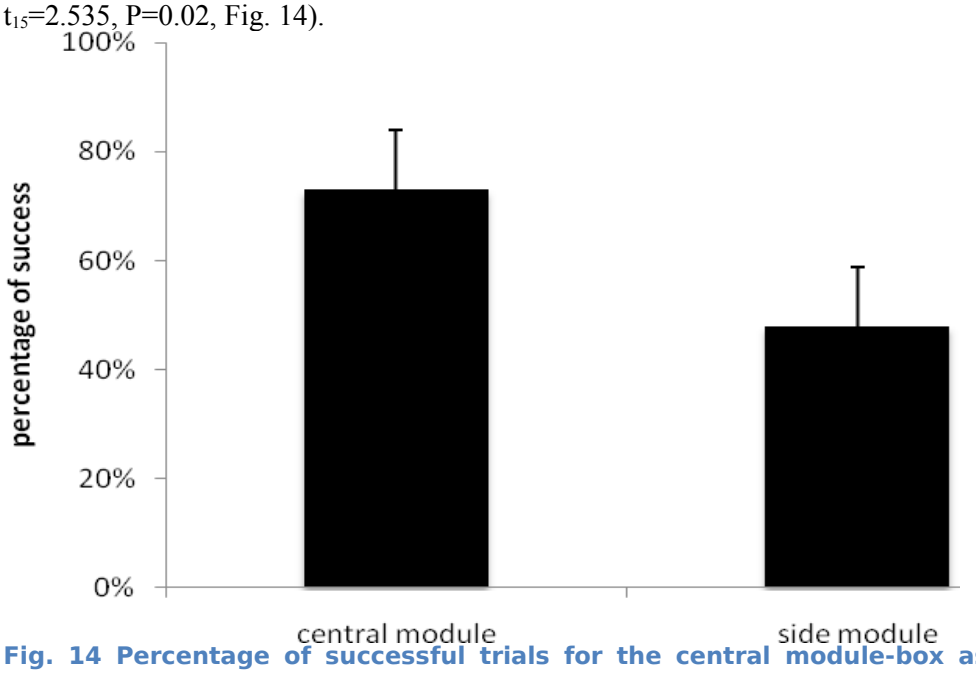


Fig. 14 Percentage of successful trials for the central module-box association and for the

lateral module-box association.

Errors

Another way to look at the subjects' performance is in terms of number of errors, i.e. by considering what they do before reaching success. Overall, individuals could fail to obtain the reward in three different ways. First, individuals could manipulate one of the two modules not associated to the rewarding box ('Module error'). Second, individuals could manipulate the 'correct' module but fail to produce the action that is associated to the rewarding box ('Action error'). Third, individuals could fail because they act on the 'correct' module and perform the correct action but not efficiently enough to produce the opening of the rewarding box ('Efficiency error').

Results showed that the number of errors was not influenced either by the group (LMM on 180 data-points: z=-0.68, P=0.497) or by condition (LMM: z=0.67, P=0.501). The frequency with which individuals opened (Experimental group) or lighted (Control group) each box during Phase 1 did not influence the number of errors in Phase 2 (LMM on 60 data-points: left box: z=-0.48, P=0.633, central box: z=0.60, P=0.549; right box: z=0.20, P=0.838).

Fig. 15 reports the mean number of the different types of errors per individual. As shown by the graph, there was no group's difference in the number of errors performed (Anova, main effect of Group: F_1=0.498, P=0.493). The type of errors differs significantly across them (Anova, main effect of Errors: F_2=26.715, P<0.001). Indeed, the Module Error was the most representative type of error, differing significantly from the Action (post hoc test: P<0.001) and Efficiency Error (post hoc test: P<0.001).

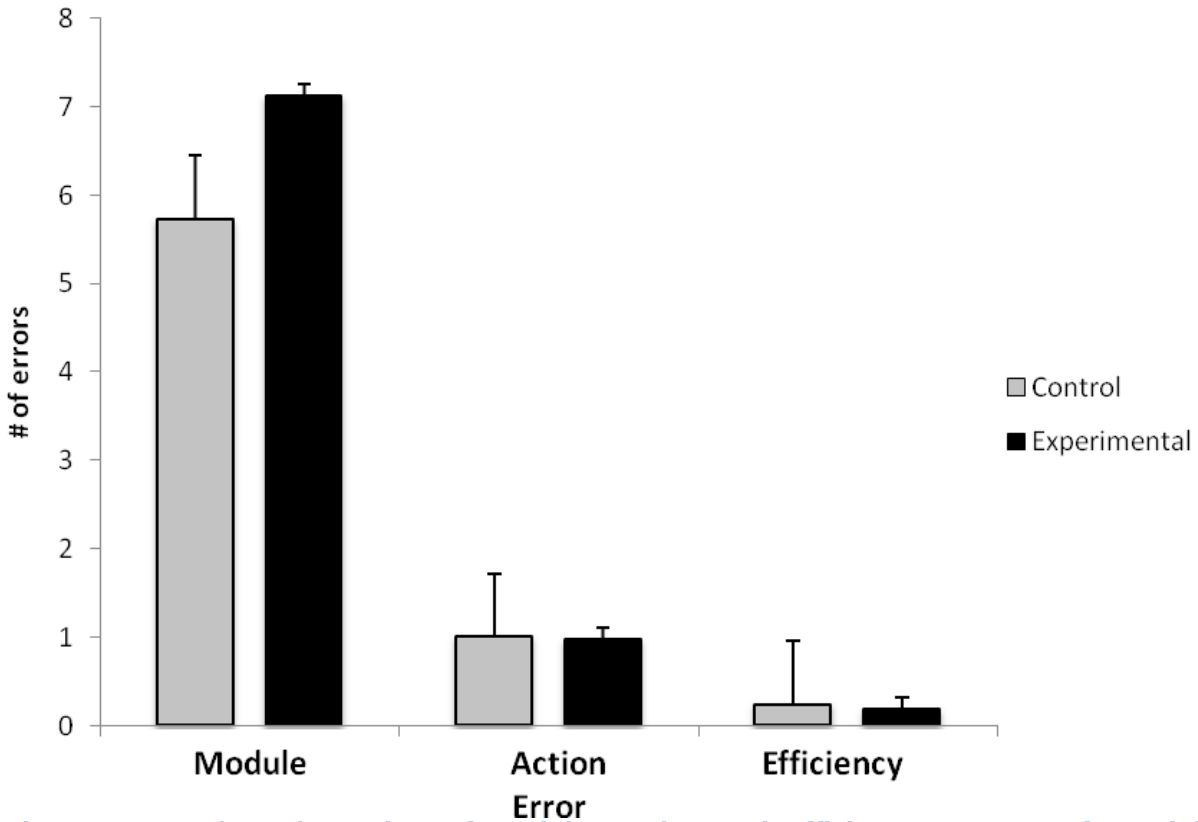


Fig. 15 Mean (+ SE) number of Module, Action and Efficiency errors performed by the experimental and control group.

Here, we focus our analyses on the Module errors. In particular, we found that when the subject performed a Module error, the choice between the two incorrect modules was not at chance level. Subjects were significantly biased towards the incorrect module located in front of the rewarded box (t_{15}=2.992, P=0.0035, Fig. 16). Errors were thus strongly biased by the reward position.

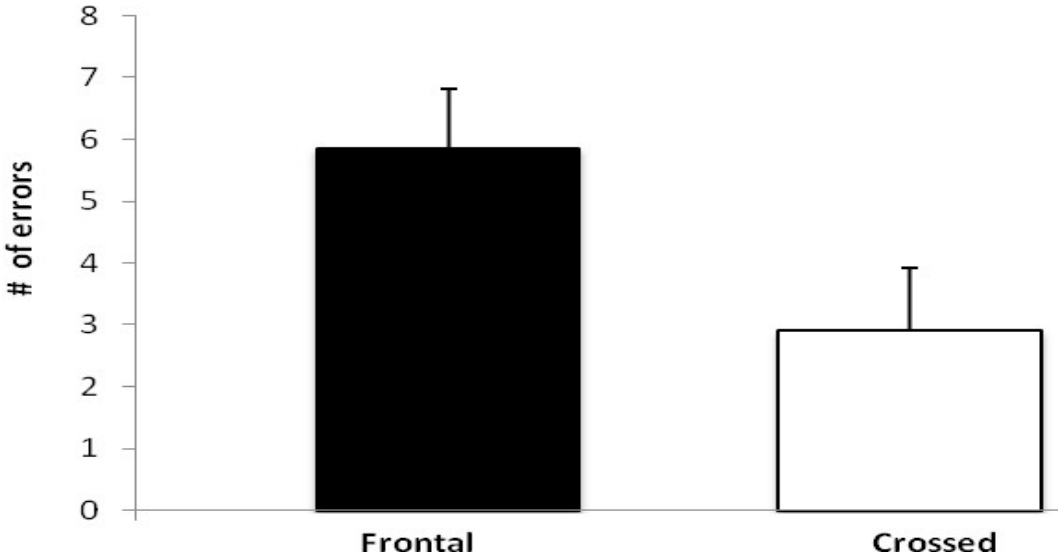


Fig. 16 Mean (± SE) number of Module errors performed on the incorrect module located in front of the rewarded box (Frontal) and on the other incorrect module (Crossed).

Spatial biases

Results showed that overall the mean number of errors produced while the subject had to open the central box (i.e. had to operate the right or left module) was significantly higher than those produced when the subject had to open the right (LMM on 180 data-points: z=-2.84, P=0.005) and left box (LMM: z=-3.38, P=0.001).

Frontal vs crossed associations

For those boxes for which the association with modules varied across conditions (i.e., the right and left boxes), there was no difference in the number of errors performed in the frontal and crossed module-box association (LMM on 60 data-points left box z=-1.11, P=0.265; right box z=-0.64, P=0.524, Fig. 17). The performance on the central box was not analyzed because in this case, the box-module association was crossed in both conditions.

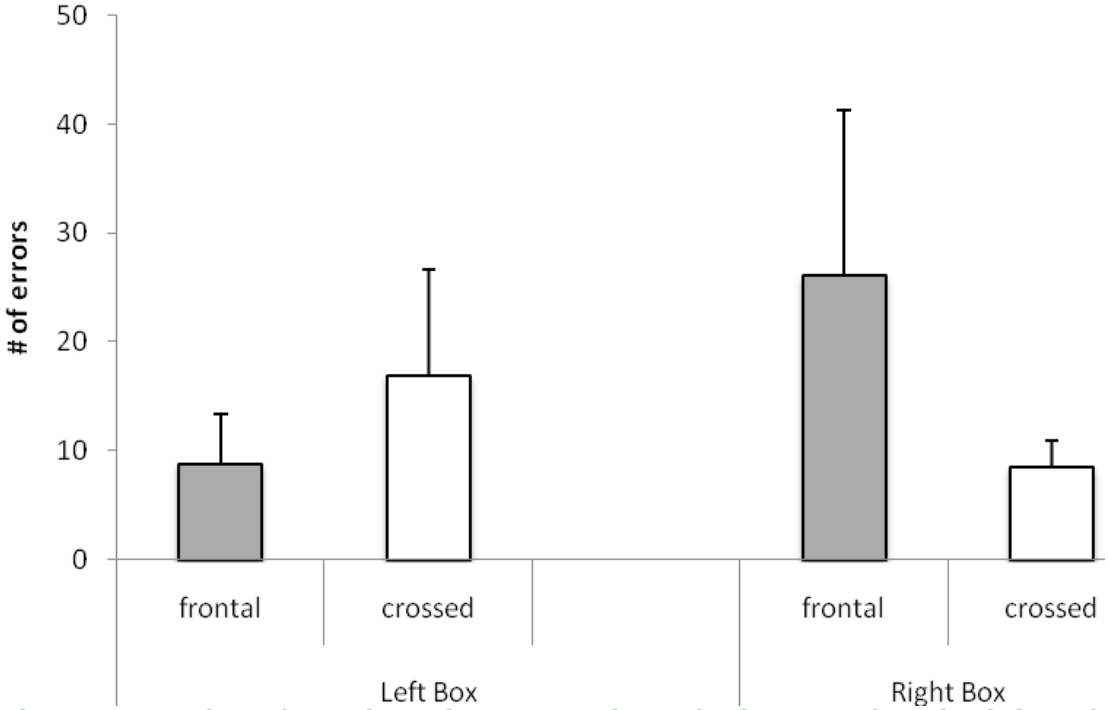


Fig. 17 Mean (+ SE) number of errors performed when opening the left and right box by operating on a frontal or a crossed module.

Number of affordances and performance

When the module-box associations were frontal, the number of errors on the left box (associated with the two affordance-module; Mean ± SE 3.6 ± 1.41) tended to be lower than that on the right box (associated with the three affordance-module; Mean ± SE: 9.6 ± 3.7), but the difference did not reach statistical significance (t-test: t_{15}=-1.5, P= 0.1557, Fig. 18).

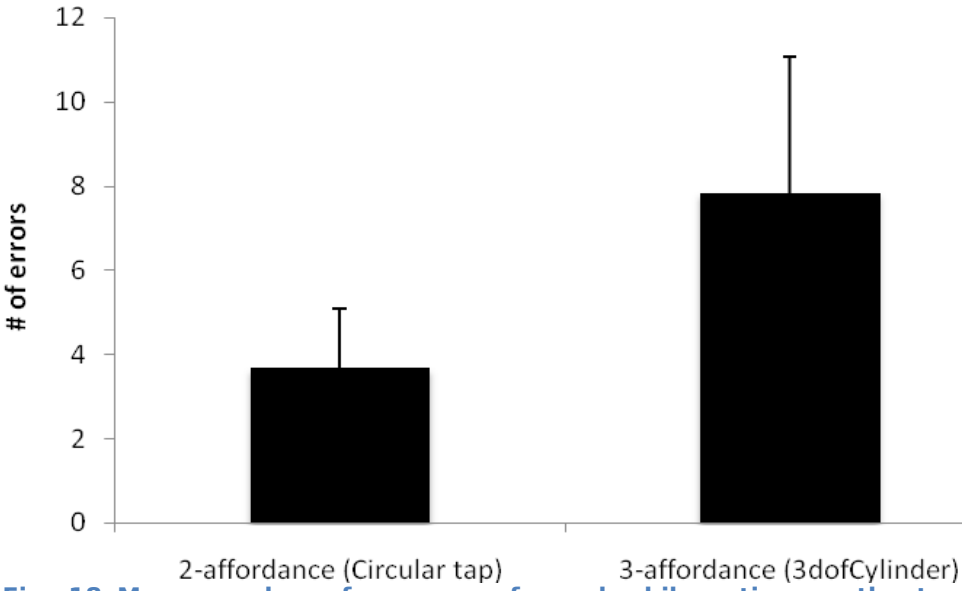


Fig. 18 Mean number of errors performed while acting on the two- and three-affordance modules.

Discussion

In Phase 1, subjects belonging to the experimental and control group did not differ in terms of latency to contact the board and general arousal. The results of modules manipulation and of boxes exploration indicate a tendency of the experimental group to act more on them than the control group, suggesting a higher level of interest when boxes could be opened than when only illuminated; these differences did not reach statistical significance probably due to the high variability among subjects.

In the 2 two-affordance modules (Circular Tap and Fixed Prism) subjects did not perform a given action more than the other. In contrast, in the three-affordance module (3dofCylinder) the push-pull action was significantly the least performed. Indeed, only 4 individuals (3 males and 1 female) out of 16 performed the action. On the basis of our observations, we suggest that the component of the 3dofCylinder linked to this action was probably not salient enough to elicit individuals' interest.

In Phase 2, the success (number of times a reward was obtained) did not differ between experimental and control group, however for the experimental group there was a tendency towards a positive correlation between the time of modules manipulation during Phase 1 and the percentage of success. This result might indicate that a higher exploration of the board increases the opportunity to solve the subsequent problem-solving task.

Overall, the box-module association (frontal or crossed) did not affect individuals' success (in terms of percentage of rewarded obtained). Although this result suggests that monkeys apparently perceive crossed associations as challenging as frontal ones, further analyses indicate that this might not be the case. Indeed, a closer look to the crossed associations of condition 1 and 2 (see Fig 4 of experimental conditions in the Methods section) reveals that monkeys obtained a higher number of rewards from the lateral boxes than from to the central one, since the lateral boxes were under the control of the central module that was strongly preferred. Therefore, the results between frontal and crossed box-module associations could be due to this bias and not by the difference in associations per se.

The type of errors more frequently performed by both experimental and control groups was due to actions on incorrect modules. These 'Module' Errors were generally directed towards the module placed in front of the rewarded box (frontal error). This result, along with the higher number of rewards obtained by operating the central module (i.e., linked to lateral boxes) explains why we found a significant higher number of errors for the central box, that in both conditions required actions from the lateral modules.

Modules varied for the number of affordances. The results showed that there was a tendency to perform less error when manipulating a two-affordance module (i.e., Circular Tap) compare to the three-affordance ones (3dofCylinder), suggesting that the complexity of the modules could negatively influence individuals' performance. However, this result may also be due to the fact that the Circular Tap was manipulated less than the 3dofCylinder, reducing in turn the opportunity to perform errors.

In conclusion, our data only partly support the prediction that subjects exploring more during Phase 1 should perform better in the Phase 2. In contrast, there is no support for the prediction that objects presenting more affordances are more explored and that the experimental group performs better than Control group. It is possible that the age of the subjects (i.e., they were all fully adults) negatively affected individuals' interest towards manipulanda leaving little space for learning new actions and/or novel objects affordances. Moreover, the monkeys were very accustomed to be rewarded for what they do in the testing room, having participated to a large variety of experiments involving food. This is another factor that might have lowered their intrinsic motivation to explore modules for its own sake.

Results with Children (UCBM)

Subjects

Experiment with pushbuttons

Six children aged between 24 and 51 months were involved in the experiment with pushbuttons. 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).

- Age group 1, mean age: 24 months
- Age group 2, mean age: 31.5 months
- Age group 3, mean age: 49.5 months

SUBJECT	GROUP	AGE [months]
CBM08	EXP	24
CBMO6	CTRL	24
CBM11	EXP	33
CBM09	CTRL	32
CBM17	EXP	48
CBM19	CTRL	51

TABLE 1 Subjects involved in the experiment with pushbuttons. Children were age-matched and assigned to one of two groups: EXPerimental and ConTRoL

Results

Results of the push buttons protocol

During training phase both CTRL and EXP groups were exposed to the board for 10 minutes. In the CTRL group, box openings were disabled (see section 1.2). Table 2 summarizes the interaction of each subject with the board during this phase.

SUBJECT	AGE [months]	GROUP	Pushes	Correct actions*	activation box 1 (%)	activation box 2 (%)	activation box 3 (%)
CBM08	24	EXP	142	57	14.03	42.11	44.86
CBMO6	24	CTRL	292	27	18.51	33.33	48.16
CBM11	33	EXP	92	19	21.05	36.84	42.11
CBM09	32	CTRL	102	59	25.42	30.51	44.07
CBM17	48	EXP	239	49	12.24	48.98	38.78
CBM19	51	CTRL	365	36	36.11	30.56	33.33

TABLE 2 Number of interactions of each subject with the board during training phase. * Refers to the percentage of button pushes lasting longer than 2s.

Younger subjects (Age group I and II) seem to prefer middle and right pushbuttons (respectively RB and GB) possibly because subjects are right-handed. Such a preference is not observed in age group III children. Interestingly, there is a progressive increase of the number of pushes of the LB from the younger to older age. (see Fig. 19)

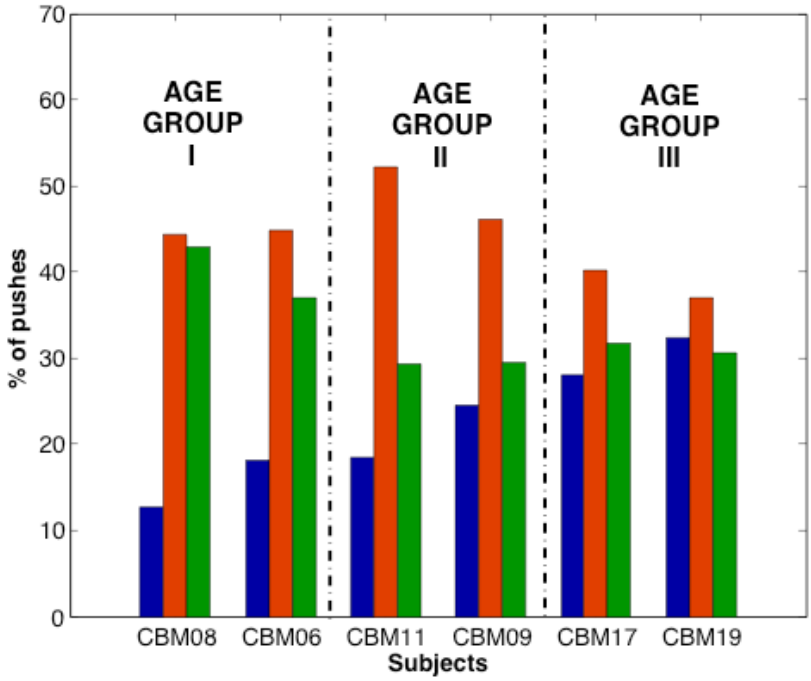


Fig. 19 Percentage of pushes for each button: blue, red, and green

No significant differences in terms of number of correct actions were observed between the experimental and control groups.

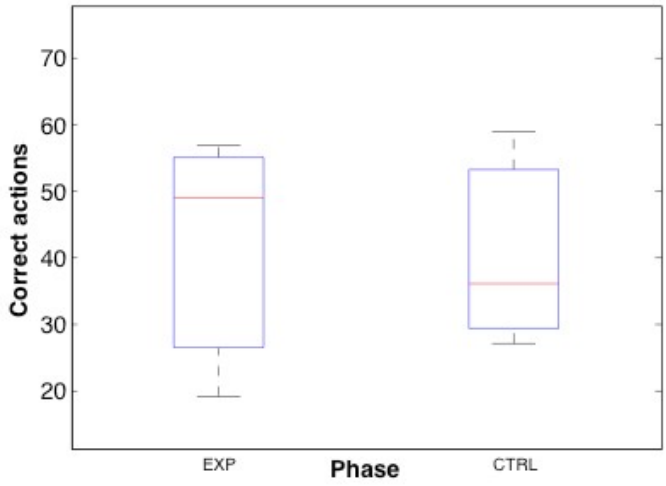


Fig. 20 Number of correct actions: Comparison between CTRL and EXP group

To assess the transfer of motor skills into the new context, during Test Phase subjects were asked to retrieve a sticker inserted into one of the three boxes. Nine stickers were used (three for each box), the insertion order was random (see section 2.2 for more details). Subjects in the experimental group were found to retrieve retrieve a higher number of rewards. The training effect increases dramatically with age (fig. 10).

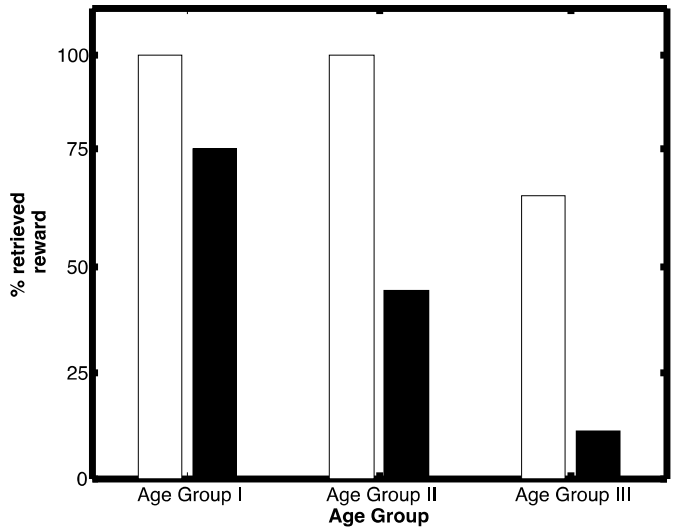


Fig. 21: Retrieved rewards: comparison between EXP (withe) and CTRL (black) group

The differences between EXP and CTRL groups can be mainly ascribed to the buttons with crossed relationship with the boxes (Cb, Rb, see Fig. 7): CTRL subjects are not able to retrieve the reward inserted in such boxes during the test phase, although the corresponding buttons are those more frequently explored during the training phase.

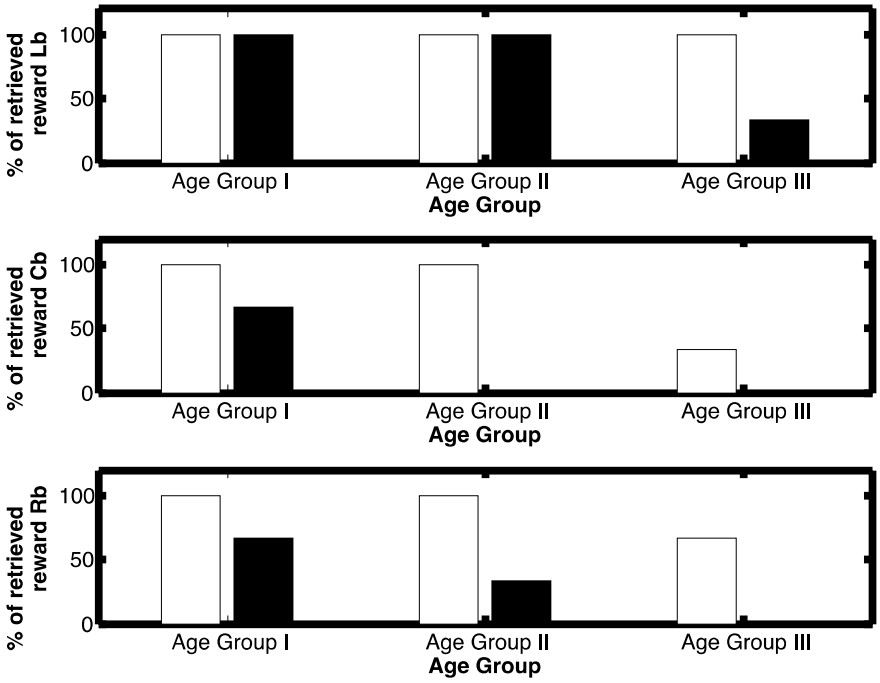


Fig. 22: Percentage of retrieved rewards for each box: Left box (Lb), Central box (Cb), and Right box (Rb). White bars: EXP group; black bars CTRL group

To assess if subjects have learned the spatial relationship between buttons and boxes we defined a Spatial Relationship Index (SRI) as:

$$SRI = \frac{n^\circ \, of correct \quad pushes}{n^\circ \text{ of total pushes per trial}}$$

If a subject pushes only that button which controls the opening of the box where the reward is placed, such index will tend to one; if a subject pushes randomly all the buttons such index will tend to 0.33; if a subject learns a wrong relationship such index will tend to zero.
In Fig. 23, the boxplot of SRI for the six subjects involved in the experiment is represented. Red lines represent the median value of the index. Younger subjects (Age Group I) do not seem to have learnt the spatial relationship between buttons and boxes; in Age Group II, the two medians seem suggest that could be a difference between CTRL and 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

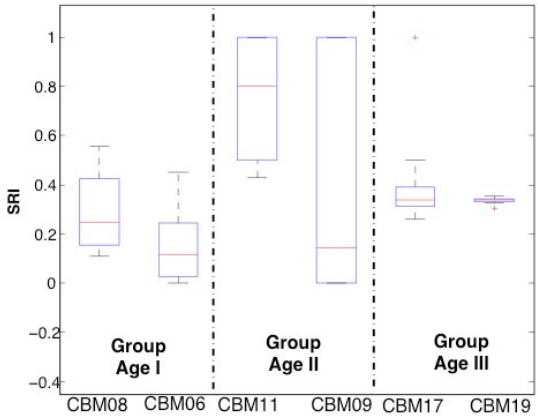


Fig. 23 Spatial Relationship Index measured for each subject

EXP group have learned that pushing a button for a longer period opens the boxes and are able to transfer this motor skill to a new context, in order to retrieve a reward. However, they did not discover the spatial relationship between the buttons and the boxes.
Considering separately that trials where reward was placed in boxes cross related with pushbuttons (Cb and Rb), children of Age Group II seem to show a statistically significant difference (*F(1,10)=25.72, p=0.0007) in spatial Relationship index between CTRL and EXP.

Subjects	SRI simple [MEAN ± SD]	SRI crossed [MEAN ± SD]
CBM08	0.39 ± 0.24	0.24 ± 0.10
CBMO6	0.20 ± 0.22	0.14 ± 0.17
CBM11	0.95 ± 0.1	**0.62 ± 0.23***
CBM09	1	**0.08 ± 0.12***
CBM17	0.34 ± 0.01	0.45 ± 0.28
CBM19	0.32 ± 0.01	0.33 ± 0.01

TABLE 3 Spatial Coherence Index for each subject

Discussion of the results and future work
We think that the results of the first experiments on intrinsically motivated learning and transfer of motor skills are very encouraging. We need to increase the number of subjects in the two groups, in order to ask following questions:

- effect of the position of the buttons vs. the color (color preference?)
- effect of age (do older children discover the spatial relationship between buttons and boxes?)
- furthermore, we will assess hand preference using a standard inventory

Results of the mechatronic modules protocol

Preliminary results of the pushbuttons protocol seems to be very interesting, for this reason UCBM-LDN plans to increase the number of subjects involved in the experiment with pushbuttons using subjects scheduled for the experiments with mechatronic modules. Experiments with mechatronic modules will be carried out from September on with newly recruited children using the same protocol already set for monkeys and discussed above.
UCBM-LDN decided to follow this strategy after a meeting with ISTC-LOCEN and ISTC-UCP (April 19, 2011), where preliminary results were discussed and considered to be a suitable input for the neural network control architecture of iCub robot for the the CLeVeR-B demonstrator.

Design of new experiments with children

To assess cumulative learning in children we are planning to perform a second set of experiments. The main issue to address is about the way in which children link together simple skills to develop more complex behaviours. The learnt skills could be both motor and cognitive: spatial relation between buttons and boxes in the experiment with pushbuttons for example, is a cognitive skill while the ability to keep pressed a button for at least 2 seconds is an example of motor skill.

To investigate how these skills are linked together, several experiments were proposed that should be discussed at the review meeting in July.

We are considering two kinds of experiments:

1. Motor sequential experiment

Bimanual coordination is a fundamental motor skill that develops relatively late. Bimanual coordination can have different degrees of difficulty, reaching from simple tasks such as unscrewing a bottle to complex tasks like playing two different rhythms with both hands. Learning of complex bimanual coordination tasks is sequential, as it involves learning the movements of each hand separately, and then putting both hands together. We will assess sequential learning of bimanual coordination with the pushbutton platform, implementing a protocol where the opening of the box is enabled only if the child presses one button for more than 2 sec while holding the other button pressed for the previous 2 sec. The experiment has the same 2 phases structure of the first set of experiment: a learning phase, during which the child freely explore the board to learn how the action he/she perform modify the state of the board; a test phase, during which the experimenter assesses how much child has learnt. The children will be divided into an experimental group and in a control group as in the previous protocol.

2. Cognitive combined experiment

The experiment has the same 2 phases structure of the first set of experiment: a learning phase, during which the child freely explore the board to learn how the action he/she perform modify the state of the board; a test phase, during which the experimenter assesses how much child has learnt. In the learning phase child is exposed to the board for 10 minutes. The board is programmed to turn on a green light when child performs a specific action on one of the three modules plugged in. The green light remain turned on for at least 5 seconds after the end of the action. The other two modules could return audio and video feedback to the child but the correct action is rewarded only if the green light is switched on. In the test phase nine stickers are randomly put inside the box and child is asked to retrieve them.

Design of new experiments with monkeys

To design the next experiment with the monkeys we need the results of the experiment described above. Since all subjects are now familiar with the board as a source of (extrinsic) reward, further experiments on intrinsic motivation with the same board are unfeasible. A possibility is that future experiments will explore implicit knowledge by exploiting the richness/redundancy of the board feedbacks (proprioceptive feedbacks, audio and visual effects) to assess how they affect learning processes.

The results of the above experiment suggest that although free exploration may improve individuals' success in a subsequent problem-solving task, monkeys were not able to learn form their previous experience with the board. Several factors may account for this. First, all our subjects were adults and this could have decreased their propensity to explore and to learn without extrinsic rewards. Second, the strong tendency to occupy the centre of the board led to higher investigation rates directed to the central manipulandum and, consequently to higher rates of opening the associated lateral box, regardless of whether this was or not the rewarded box. This bias also affected the likelihood to solve crossed and frontal associations. Third, during Phase 1, monkeys explored the boxes but with little interest since they were not rewarded. This could have decreased their attention and their ability to recall action-outcome associations during Phase 2.

On this basis, we aim to improve our experimental set-up so as to clarify the role of intrinsic motivation on monkeys' performance and plan a new set of experiments that will investigate monkeys'

cumulative learning processes with experimental procedures that seem particularly promising for addressing this topic of the project.

I. Improvements to the experiment on intrinsic motivation in monkeys

The set up will be the same of the first generation of experiments, but we will improve the design in order to strengthen the interpretation of the results already obtained. First, we will add two juvenile individuals to our sample in order to have a better insight on whether age affects performance. Second, in order to overcome the bias toward the central position we will use two manipulanda, one at each side of the board, while leaving the central position empty. Third, to increase subjects' interest toward the boxes, we will insert an object inside them; the objects will be chosen so that they will not be perceived as food.

As in the previous experiment, subjects will be assigned to the Experimental and the Control group. Both groups will undergo a first free exploration phase (Phase 1) and then will be tested in problem-solving tasks (Phase 2) that will require the ability to recall the action-outcome association (module-box association) learned in Phase 1.

II. Experiments on cumulative learning in monkeys

Here below we provide two alternative experiments.

Experiment 1: Cumulative learning processes in tool use

Capuchin monkeys efficiently detect objects' properties. Visalberghi et al. (2009) demonstrated that they choose a stone tool to crack open nuts on the basis of visual (size) and non-visual functional clues (friability, weight). More recently, Manrique et al. (in press) reported that capuchins are highly successful in selecting tools on the basis of rigidity. In the latter experiment, subjects inferred the property of the objects observing the experimenter playing or manipulating it in the absence of an extrinsic reward, and then used it to efficiently recover an out-of-reach reward.

Given the above results, we propose to carry out an experiment on cumulative learning aimed to assess whether capuchins can learn to use a tool to obtain another tool that could be used to obtain out of reach food.

Subjects will face an out-of-reach liquid reward inside a 90° bended Plexiglas tube. This tube will be fixed and baited in one corner of the cage and a table will be in front of the other corner of the cage. The experimenter will take one tool and will show its rigidity properties (by bending and unbending it 5 times). Immediately after, the tool will be placed on the table out of subject's reach. Then the experimenter will hand another tool to the subject. Since the tool demonstrated and the tool given to the subject will be flexible or rigid, there will be 4 different conditions:

Tool demonstrated to the subject	***Tool handed to the subject***
Flexible	Rigid
Flexible	Flexible
Rigid	Flexible
Rigid	Rigid

When a flexible/correct tool is given to the subject, the reward inside the tube can recovered with it. When a rigid/wrong tool is given, the subject will not be able to reach the reward and should first recover the flexible/correct tool placed on the table. We thus expect that capuchin monkeys will learn to sequentially use the tools (first the rigid tool to get the flexible one, and then the flexible one to reach for the reward) after having explored in a previous phase (and in the absence of extrinsic rewards) the tool properties.

Experiment 2: Cumulative learning processes in concept learning

Capuchins acquire abstract concepts on the basis of perceptual equivalence between stimuli. We have recently demonstrated their ability to learn *same* and *different* concepts in a relational matching-to-sample task (Truppa et al., submitted). The relational MTS (RMTS) task requires subjects to understand whether the relationship among attributes of objects belonging to one set is equivalent to the relationship among other objects belonging to another set. This task is very difficult and only one individual succeeded. The detailed analysis of the complete sequence of the training sessions concerning the successful subject evidenced a very distinctive learning trend. Initially, the subject seemed to spontaneously decompose the task in two sub-problems. Its learning pattern reveals that first it reached criterion on the same trials and then it reached criterion on the different trials. Its success in the different condition co-occurred with a worsening of performance on the same condition. Eventually, in the last part of the learning process, it recombined the knowledge previously acquired separately becoming concurrently successful in both conditions. On this basis Truppa and collaborators argued that to learn the *sameness* concept and the *difference* concept "at once" is very demanding in terms of attentive resources and/or working memory load.

Here we propose to carry out an experiment in which subjects will receive the "same" trials and the "different" trials separately and compare their performance with that observed when the two types of trials are intermixed. Our prediction is that presenting *same* and *different* conditions separately would make the task easier than presenting both conditions at once, since the latter requires more parallel processing of information.

The experiments described above on intrinsic motivation and cumulative learning can be run by our team starting from November 2011 (i.e., after monkeys and children data will be systematically compared) and during 2012. Both experiments on cumulative learning are scientifically challenging and provide testable hypothesis. However, given the time constraints they pose in terms of number of daily sessions, we will carry out only one of the two. Our decision regarding the two proposed experiments will be discussed during the next project meeting so to ensure the general consensus among project partners, and especially CNR-ISTC-LOCEN, UCBM and USFD.

REFERENCES

Balleine, B. W., & O'Doherty, J. P. (2009). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. Neuropsychopharmacology, 35(1), 48-69.

Bednark, J.G., Reynolds, J.N.J., Stafford, T., Redgrave, P. and Franz, E.A. (submitted). Electrophysiological correlates of acquiring action-effect associations.

Bompas, A., & Sumner, P. (2008). Sensory sluggishness dissociates saccadic, manual, and perceptual responses: An S-cone study. *Journal of Vision, 8*(8).

Dickinson, A. (1980). Contemporary animal learning theory. Cambridge Univ Pr.

Fragaszy, D., Visalberghi, E. & Fedigan, L. (2004). *The complete capuchin: the biology of the genus Cebus*, Cambridge University Press

Glow, P. H., Roberts, J. E., Russell, A. (1972). Sound and light preference behaviour in naïve adult rats. Australian Journal of Psychology, 24, 173-178.

Glow, P. H., Winefield, A. H. (1978). Response-contingent sensory change in a causally structured environment. Animal Learning and Behaviour, 6, 1-18.

Harlow, H. F. (1950). Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. Journal of Comparative and Physiological Psychology, 289-294

Harlow, H. F., Harlow, M. K., & Meyer, D. R. (1950). Learning motivated by a manipulation drive. Journal of Experimental Psychology, 40, 228-234.

Jacobs, GH (1998) A perspective on colour vision in platyrrhine monkeys. Vision Research, 38, 3307-3313.

Klemke E.D., Hollinger R., Kline A.D., Introductory Readings in the Philosophy of Science. Prometheus Books, New York, 1980 .

Manrique HM, Sabbatini G, Call J, Visalberghi E (in press) Tool choice on the basis of rigidity in capuchin monkeys. Animal Cognition

Martin P., Bateson P., Measuring Behaviour: an introductory guide, 3rd ed., Cambridge University Press, Cambridge , 1998.

Redgrave, P., Gurney, K., & Reynolds, J. (2008). What is reinforced by phasic dopamine signals? Brain Res Rev, 58(2), 322-39.

Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M. C., Lehericy, S., Bergman, H., Agid, Y., et al. (2010). Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. Nature Reviews. Neuroscience, 11(11), 760-772. doi:10.1038/nrn2915

Redgrave, P., Gurney, K.N., Stafford, T. and Thirkettle, M.(2010). The Role of the Basal Ganglia in Discovering Novel Actions. In Baldassarre G., Mirolli M. (eds.) (2010). Roadmap Book. Deliverable D7.1 of the EU-funded Integrated Project"IM-CLeVeR - Intrinsically Motivated Cumulative Learning Versatile Robots"

Stafford, T., Walton, T., Hetherington, L., Thirkettle, M., Gurney, K. N., & Regrave, P. (2010). A novel behavioural task for researching intrinsic motivation. In Baldassarre G., Mirolli M. (eds.) (2010). Roadmap Book. Deliverable D7.1 of the EU-funded Integrated Project"IM-CLeVeR - Intrinsically Motivated Cumulative Learning Versatile Robots"

Taffoni, F., Vespignani M., Formica D., Cavallo G., Polizzi di Sorrentino E., Sabbatini G., Truppa V., Visalberghi E., Keller F., Guglielminelli E., (in preparation) A mechatronic platform for behavioral analysis of nonhuman primates. Journal of Integrative Neuroscience

Truppa, V., Piano Mortari, E., Garofoli, D., Privitera, S., Visalberghi, E. (submitted). Same/different concept learning by capuchin monkeys in matching-to-sample tasks.

Visalberghi E, Addessi E, Truppa V, Spagnoletti N, Ottoni E, Izar P, Fragaszy D. (2009) Selection of effective stone tools by wild bearded capuchin monkeys. *Current Biology* **19**, 213-217.

Wasserman E.A., Zentall T.R. Comparative Cognition: Experimental Explorations of An-

imal Intelligence, Oxford University Press, New York , 2006.

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., Knowlton, B. J., & Balleine, B. W (2006). Reversible inactivation of dorsolateral striatum enhances sensitivity to changes in action-outcome contingency in instrumental conditioning. Behavioural Brain Research, 66(2), 189-196