4th International Conference on
Development and Learning and on Epigenetic Robotics
October 13-16, 2014. Palazzo Ducale, Genoa, Italy

Learning where to Look with Movement-Based Intrinsic Motivations: A Bio-Inspired Model

Valerio Sperati
Consiglio Nazionale delle Ricerche
Istituto di Scienze e Tecnologie della Cognizione
Laboratory of Computational Embodied Neuroscience
(CNR-ISTC-LOCEN), Roma, Italy
Email: valerio.sperati@istc.cnr.it

Gianluca Baldassarre
Consiglio Nazionale delle Ricerche
Istituto di Scienze e Tecnologie della Cognizione
Laboratory of Computational Embodied Neuroscience
(CNR-ISTC-LOCEN), Roma, Italy
Email: gianluca.baldassarre@istc.cnr.it

***Abstract*—Most sophisticated mammals, in particular primates, interact with the world to acquire knowledge and skills later exploitable to obtain biologically relevant resources. These interactions are driven by intrinsic motivations. Recent research on brain is revealing the system of neural structures, pivoting on superior colliculus, underlying trial-and-error learning processes guided by movement-detection, one important element of one specific type of intrinsic motivation mechanism. Here we present a preliminary computational model of such system guiding the acquisition of overt attentional skills. The model is formed by bottom-up attentional components, exploiting the intrinsic properties of the scene, and top-down attentional components, learning under the guidance of movement-based intrinsic motivation. The model is tested with a simple task, inspired by the 'gaze-contingency paradigm' proposed in cognitive psychology, where looking some portions of the environment can directly change it. The tests of the model show how its integrated components can learn skills causing relevant changes in the environment while ignoring changes non-contingent to own action. The model also allows the presentation of a wider research agenda directed to build biologically plausible models of the interaction between overt attention control and intrinsic motivations.**

I. Introduction

Motivations play at least two major functions in animal adaptation. First, they support the selection of useful behaviours. Second, and relevant for this work, they generate learning signals. *Intrinsic motivations* (IMs) [1], contrary to *extrinsic motivations* (EMs), drive more sophisticated animals, in particular primates, to acquire skills and knowledge independently of the achievement of resources directly related to biological fitness such as water, food, and predator avoidance [2]. IM mechanisms generate learning signals by directly monitoring the level or improvement rate of the skills and knowledge to be acquired [3], rather than referring to homeostatic body processes related to the attainment of useful material resources as EMs tend to do [2].

The literature on IMs is showing that there are different types of IM mechanisms [3], [4], [5]. The most studied are: (b) *Prediction-based IMs*: these are triggered by stimuli that violate the animal's expectations [6], [7], [8]; (a) *Novelty-based IMs*: these are triggered by stimuli that are not in the animal's memory [5]; (c) *Competence-based IMs*: these are related to the animal's capacity to change the world with its actions [3], [9], [10], [11], [12]. These mechanisms can interact in various ways to support skill and knowledge acquisition. An example of this might be *agency*, defined by some authors as the "agent's sense that it is the cause or author of some effects" [13]. Indeed agency, shown to have a potent role in motivating spontaneous interactions with the environment in children and monkeys [14], [15], possibly involves surprise, to detect environment changes, and competence-related mechanisms, linking those changes to own action. This work might contribute to clarify these phenomena.

Neuroscientific research is starting to reveal specific brain mechanisms underlying IMs. Relevant for the model presented here and for the study of agency, Redgrave and colleagues [16] have proposed a theory on *superior colliculus* (SC), an important component of the vertebrate midbrain, related to prediction-based IMs and their guidance of competence improvement. The superficial layers of SC receive input from eyes and higher levels of the visual cortex and based on this its deeper layers play a key role in eye movements. Along with these sensorimotor functions, the SC also seems to play important functions for motivation and learning. In particular, SC strongly reacts to luminance changes and on this basis triggers the production of phasic bursts of *dopamine*. Dopamine is a fundamental neuromodulator guiding the trial-and-error learning processes of *basal ganglia-cortical loops* [17], [18]. The idea is that SC can detect possible effects (changes) that the animal actions produce on the environment and on this basis drive the improvement of the skill that causes them. The skills so acquired can be later exploited when the effects they can produce become desirable [19], [20], [21]. The learning signal produced by SC fades away when, with experience, the action effect becomes more predictable for the animal [16]. SC movement detection can thus be related to prediction-based IMs and competence acquisition [5].

Here we focus on the IM-driven acquisition of *overt attention skills* related to the capacity of the agent to direct the eye gaze on relevant portions of the scene so as to collect useful information from it. We propose a model of these processes based on some principles playing an important role in primate attention [22]: (a) the fovea-periphery organisation of the eye, and the active control of the eye gaze; (b) a bottom-

up attention process guiding visual exploration on the basis of image 'objective' features (e.g., high contrast, movement, etc.); (c) a top-down attention process acquiring attentional skills on the basis of trial-and-error processes guided by the agent's needs and goals (here represented by IM rewards).

The model is part of a broader research agenda directed to model how prediction-based and novelty-based IMs allow the acquisition and later exploitation of overt attention skills, and how these serve manipulation. This agenda involves modelling and studying: (a) bottom-up and top-down attention and their interactions [22], [23], [24]; (b) prediction-based (this work) and novelty-based IM mechanisms generating learning signals, and how they guide top-down attention learning; (c) the interplay between attention and manipulation behaviours and underlying processes [22].

We previously modelled and studied the interaction of bottom-up and top-down attention processes [22], [23], [24] but we did not relate this to biologically-plausible mechanisms for IM-based learning guidance. The model presented here shares with those previous models the importance given to the interplay between bottom-up and top-down attention. However, it also presents the following innovations: (a) it has a system-level architecture following some main aspects of the corresponding brain structures and functions; (b) it reproduces the SC-like generation of movement-based IM learning signals; (c) although we use simple colour-blob objects as in the previous models, here we start to introduce a simple feature-based object recognition component to test the model robustness to feature-based object representations; (d) a strong biologically-plausible reflex to look where movement happens; (e) the capacity to work without hardwired trials. To the best of our knowledge, there is no computational model integrating these features.

The model is tested with a setup inspired by the cognitive psychology experiment presented in [25]. This experiment uses the *gaze-contingency paradigm*[1] [26] in which participants can change the image of a computer screen by simply looking at some parts of it. This paradigm is ideal for studying how looking actions can drive learning processes based on environment changes. In the psychological experiment presented in [25] participants learn to direct their gaze on button-like pictures on a computer screen and this causes the sudden appearance of other elements in the screen. Here we test our model with a setup inspired by these experiments with the *goal of understanding if and how the SC detection of luminance changes can lead the agent to acquire skills that produce effects on the world while ignoring other changes happening independently of the agent's actions*.

The rest of the paper is organised as follows. Section II presents the setup used to test the model and the model itself. Section III presents the results of the tests of the model. Finally, Section IV draws the conclusions and highlights the specific planned future enhancements of the model within the research agenda introduced above.

[1]http://en.wikipedia.org/wiki/Gaze-contingency_paradigm

II. Methods

A. Simulated environment and task

The environment is formed by six spheres coloured respectively in red, green, blue, cyan, yellow and magenta (Fig. 1). All spheres are static and characterised by the same dimensions and luminance ($L = 0.5$, where 0.0 is full dark and 1.0 is maximum luminance). The spheres are spatially arranged on a black-background plane in front of the system. The setup also involves two additional white spheres ($L = 1.0$), which we call *random light* and *causal light* (see Fig. 2). The random light appears randomly, with a probability $p = 0.005$ per step, always at the same position between the magenta and red spheres, and then gradually fades away in 10 steps. The causal light appears the time step after the system looks at the blue sphere with its fovea, always at the same position between the blue and yellow spheres, and fades away in 10 steps. We refer to the blue sphere as the *button*, and to the light appearance as 'light switching on'. The environment was simulated with *OpenGL* libraries[2].

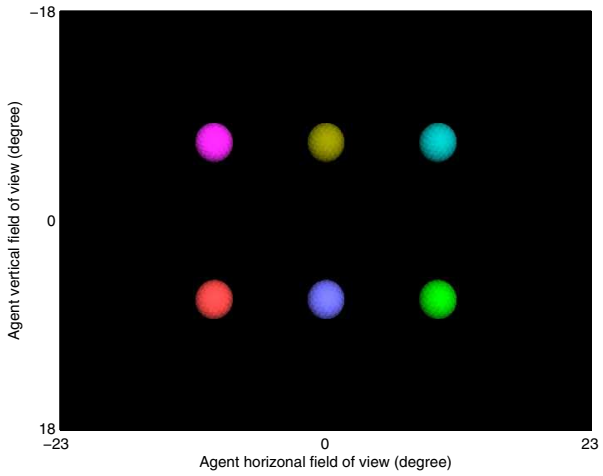


Fig. 1: Image perceived by the system when looking ahead (pan and tilt angles are set to $0°$). The environment is formed by six spheres, each one coloured differently (from top left and clockwise: purple, yellow, cyan, green, blue, red).

B. Simulated camera

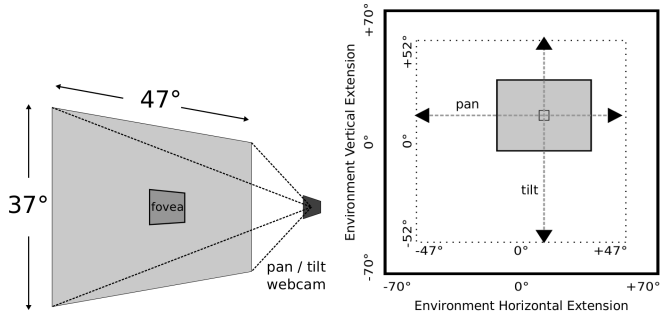
The system is based on a simulation of a real motorised camera system formed by two connected servos attached to a webcam (see Fig. 3a). Preliminary tests of the model with the real system are being carried out now so it will not be further discussed here. The camera can perform pan and tilt movements within a range of respectively $[-47°; +47°]$ and $[-52°; +52°]$. At each simulation step, the camera captures a 640×480 pixel RGB image covering a view field of $[47°; +37°]$. Given these values, the system can visually explore an area spanning $141° \times 141°$ (Fig. 3b). Notwithstanding

[2]www.opengl.org

(a) (b)

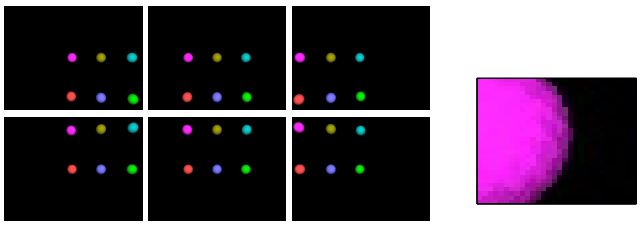
Fig. 2: (a) Appearance of the causal light (white sphere) in the step after the system looks at the button (blue sphere). (b) Random light appearance (happening with probability $p = 0.005$ per step). Lights are identical from the system point of view. Both snapshots have been captured when the system is looking ahead.

this ample range of movement, we will see that the model remains mainly focused on the sphere stimuli (Fig. 4a).



(a) (b)

Fig. 3: (a) The system consists of a simulated motorised webcam with pan and tilt movements: its fovea and periphery views are shown. (b) Movement range of the system, outlined by dots.



(a) (b)

Fig. 4: (a) Whole images perceived by the system when looking at the six different spheres with the fovea. Each snapshot (640×480 pixels) covers $47° \times 37°$ (see Fig. 3a) and involves different pan/tilt positions (Fig. 3b). (b) Fovea image (30×30 pixels) when the system looks at the purple sphere. Notice that, based on the bottom-up component, the system tends to foveate the edges of the spheres.

C. 'Task'

Coherently with the IM framework, there is not a specific task that the system should solve but rather the opportunity for it to freely explore the environment and learn how changing it. In particular, a relevant feature of the environment consists in the possibility of temporarily switching on the causal light as this is under the potential control of the system action. On this basis, we expected the following behaviours to emerge: (a) Looking at reasonable locations (the spheres) while ignoring non-informative areas (the background); (b) looking at a light when it appears; (c) learning the skill that can actively produce the appearance of the causal light (i.e. looking at the button), and performing it with increasing frequency; (d) no active search of the random light as no skill can be learnt to cause its appearance. Although simple, these behaviours are at the core of the model potential to exploit the movement-based IM mechanism to acquire skills *autonomously*, i.e. without external intervention.

D. The model

The model (controller of the system) is bio-inspired, meaning that it captures basic aspects of the macro-organisation and overall functioning of the brain system underlying the investigated behaviours. The model is formed by two main components, the *bottom-up component* and the *top-down component*, each based on several neural maps. Most of these maps use a gaze reference frame, coherently with the functioning of the primate visuo-motor system [27]. The maps compute the camera image (input) in parallel, and on this basis determine the next target of attention (output) (Fig. 5). When triggered, a saccade movement is executed in one step. The bottom-up component is responsible for generating *reflexive saccades* driven by contrasts or sudden movements in the scene, while the top-down component is responsible for generating *voluntary saccades* driven by the self-determined tasks of the system and the learning process related to them. Below we describe the two components in detail using the following notation for the formulas: bold upper case letters for matrices, bold lower case letters for column vectors (note that the elements of the model neural maps are represented as vectors), lower case letters for scalar variables.

1) Bottom-up component: The input of this component is a low resolution (60×80 pixels), grayscale version of the input image. We refer to this input image as *periphery*. Two filters are applied to it: the first detecting edges, the second detecting luminance changes. These two filters are used to mimic the activity of respectively the *parvocellular* and *magnocellular* ganglion cells in the primate visual system [28]. After a further subsampling to 47×37 pixels, these filters are combined in a *saliency map* [29].

The saliency map is the input to a neural network, named *neural field I*, mimicking the SC superficial layer [30], [31]. This network is formed by one layer of 47×37 leaky integrators with a temporal costant τ. The network is characterised by recurrent 'Mexican hat' connections, performing local excitation and distal inhibition, linked to the units' distance in

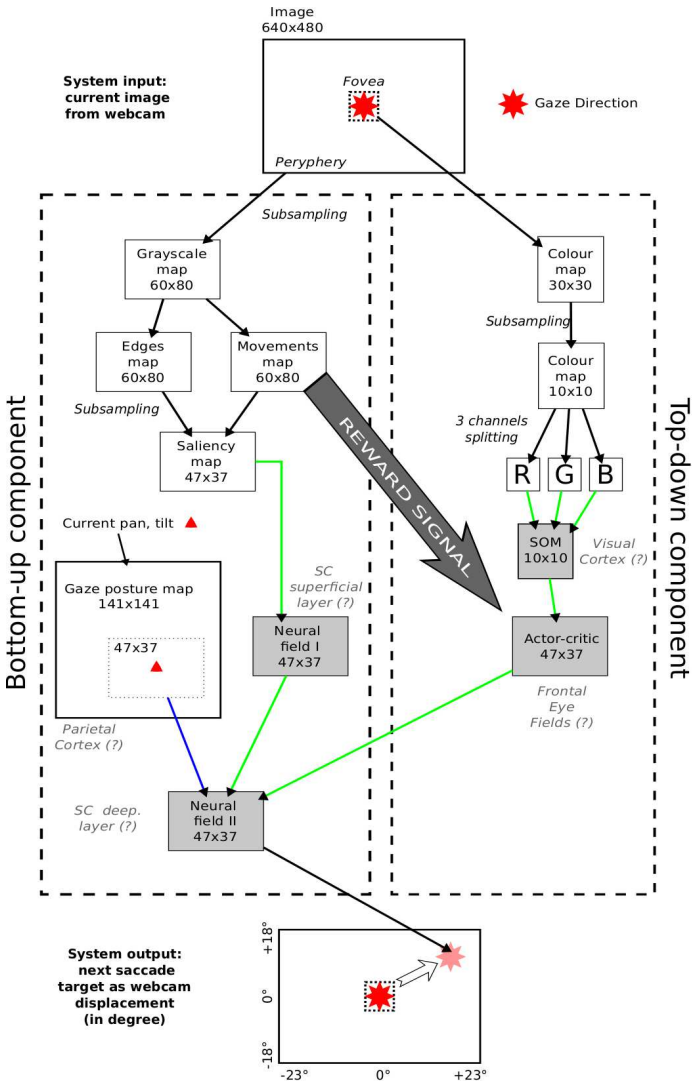


Fig. 5: Schema of the model. All maps of the model use the current gaze direction as reference frame, with the exception of the Gaze posture map using an absolute 'head' reference frame. Gray squares: maps implementing neural networks. White squares: maps implementing different visual filters. Green arrows: excitatory inputs. Blue arrows: inhibitory inputs. Light gray labels: possible brain anatomical structures corresponding to the model components. Big gray arrow: the intrinsic reinforcement signal.

the neural space [32]. This architecture leads to the formation of dynamic 'bubble-like' activities centered on most salient parts of stimuli. The activation of the map neurons is as follows:

$$\mathbf{f}_t = tanh\left(\mathbf{f}_{t-1} + \frac{1}{\tau}\cdot\left(-\mathbf{f}_{t-1} + \mathbf{i}_t + \mathbf{W}\cdot\mathbf{f}_{t-1}\right)\right) \qquad (1)$$

where $\mathbf{f}_t$ is the vector of activation of neurons at time step t, $\tau = 5$ is the temporal constant, $\mathbf{i}_t$ is the current normalised input vector (from the saliency map), $\mathbf{W}$ is the connectivity matrix (including self-connections), and $tanh$ is the 'truncated' (i.e., with negative values set to 0) hyperbolic tangent function mapping the units activations to $[0, 1]$.

Neural field I is one of three inputs of a second neural network, the *neural field II*, mimicking the behaviour of the SC deep layer [30], [31]. This is a recurrent neural network characterised by same features of neural field I except for the inhibition that is global and not proportional to the distance between neurons. This latter feature enables the evolution of only one activity bubble centred on the most salient point (winner-take-all neural competition). When a unit of neural field II exceeds a threshold θ (*winner unit*; $\theta = 0.3$), a saccade in the correspondent point of the space is triggered. The dimensions of neural field II (47×37 units) matches the field of view of the camera ($47° \times 37°$) so the winner unit position encodes the corresponding webcam desired pan and tilt angles, as relative displacements. As soon as the saccade is executed, the activity in both networks is reset, and a new activity accumulation starts.

To prevent the fixation of the system on the most salient location due to the initial bottom-up dominance, an *Inhibition Of Return* (IOR) mechanism was implemented. IOR is an attentional mechanism which promotes a better exploration of space [33]. IOR is here based on a second input from a short-term absolute spatial memory (Fig. 5), the *gaze posture map*, possibly corresponding to parietal cortex. Parietal cortex is a cortical area that processes both visual and proprioceptive information, is important for attention, and uses a body-centred reference frame in some of its components [34], [35].

The memory is implemented as a 141×141 neural network encoding the absolute space in terms of possible movements and having a Gaussian activation ($sigma = 0.5$) centred on the current absolute pan and tilt angles of the camera. The network is in particular formed by leaky units activating as follows:

$$\mathbf{m}_t = \Lambda\left(\mathbf{m}_{t-1} + \frac{1}{\tau}\cdot\left(-\mathbf{m}_{t-1} + \mathbf{p}_t\right)\right) \qquad (2)$$

where $\mathbf{m}_t$ is the vector of neurons activity at timestep t, $\tau = 240$ is a time constant, $\mathbf{p}_t$ is the current normalised input vector (Gaussian activation centred on pan/tilt values), and Λ is a function which sets to one the units which exceed this value. A submap of 47×37 units, centred on the current gaze coordinates, is cut from the gaze posture map and given to the neural field II as inhibitory input. A third input to neural field II comes from the top-down component described below.

The bottom-up component makes the system capable of exploring the visual scene by triggering saccades on the most salient points of it. When no lights are present, the system gaze jumps from a sphere to another foveating their edges. When a light appears, its high movement and contrasts overcome the other stimuli and the gaze is shifted to it. As soon as the light fades away, the attention is captured by another sphere. Note that the bottom-up component is dependent of the visual scene features and does not learn.

2) Top-down component: This component joins the information flow of the bottom-up component within the neural

field II and biases its neural dynamic toward potentially relevant points in the visual scene. The input of this component is a small (30×30 pixels) RGB patch taken from the central part of the input image. We refer to this input as *fovea* (Fig. 4b). The fovea is then subsampled to a 10×10 map and splitted in the three RGB channels. This constitutes the input for a standard 10×10 *Self Organising Map* (SOM) [32], which was pre-trained (with respect to the tests shown here) to classify the foveal input on the basis of an exploration driven by the bottom-up component. In this way the SOM learns to disambiguate the foveated stimuli (spheres and lights).

The normalised output of the SOM is used as input of an *actor-critic* neural network [36]. The actor is particular as its output is a neural map of 47×37 units which requires a special learning rule [23], [24]. The activity of the actor units is described by the following equation:

$$\mathbf{a}_t = tanh(\mathbf{W}_a \cdot \mathbf{s}_t) \tag{3}$$

where $\mathbf{a}_t$ is the vector of activities at time step t, $\mathbf{s}_t$ is the normalised input vector from the SOM, $\mathbf{W}_a$ is the actor weights matrix (subject to learning). The actor output constitutes the third excitatory input to neural field II.

The critic is a neural network with a linear output unit v whose activation is given by equation:

$$v_t = \mathbf{w}'_c \cdot \mathbf{s}_t \tag{4}$$

where v_t is the state evaluation of the critic at timestep t, $\mathbf{s}_t$ is the current normalised input vector from the SOM, $\mathbf{w}'_c$ is the transpose of the critic weight vector (subject to learning).

The learning of the actor-critic network is driven by movement: a sudden change in the visual scene when a light appears causes the production of a reward signal within the model. In particular, the reward is based on the activation of the movement map. When at least one pixel of this map exceeds a certain threshold, the reward is set to one, otherwise it is set to zero. This mimics the high sensitivity of SC to luminance changes and its capacity to generate a reward signal. In this respect, for now the reward is computed on the basis of the activation of the movement map. The reason is that making it directly dependent on the SC activation (in particular the deep layer) would cause a dopamine signal at each saccade and the biological mechanism that prevents this is still unclear (we are exploring the possibility that the response to luminance changes is inhibited within the SC when the change becomes predictable, or the alternative possibility that dopaminergic areas themselves are inhibited, cf. [16], [17]).

The *TD-error* is computed as follows:

$$\delta_t = r_t + \gamma \cdot v_t - v_{t-1} \tag{5}$$

where δ_t is the TD-error at time t, r_t is the intrinsic reward value, $\gamma = 0.95$ is a discount factor, and v_t and v_{t-1} are respectively the current and previous critic evaluations.

The critic connection weights are updated as follows:

$$\Delta \mathbf{w}_c = \eta_c \cdot \delta_t \cdot \mathbf{s}_{t-1} \tag{6}$$

where $\eta_c = 0.05$ is a learning rate, $\mathbf{s}_{t-1}$ is the previous input vector from the SOM. Importantly, note that the actor-critic learns only in the step after a saccade is performed.

The actor connection weights are updated as follows:

$$\Delta \mathbf{W}_a = \eta_a \cdot \delta_t \cdot ((\mathbf{g}_{t-1} - \mathbf{a}_{t-1}) \odot \mathbf{g}_{t-1}) \cdot \mathbf{s}'_{t-1} \tag{7}$$

where $\eta_a = 0.05$ is a learning rate, $\odot$ is the element-wise product, $\mathbf{g}_{t-1}$ is the vector containing the previous activation of neural field II, $\mathbf{a}_{t-1}$ is the previous actor output. Within the formula $(\mathbf{g}_{t-1} - \mathbf{a}_{t-1})$ implies that, with positive δ, the actor output is made closer to the neural field II output which produced a saccade, but only for the units where neural field II was active ($0 < \mathbf{g}_{t-1}$). Viceversa, for a negative δ the neural field II output is decreased. From a biological point of view, the actor captures the function of *frontal eye fields*, a cortical region in the frontal lobe playing a crucial role in generating voluntary saccades [37]. The reinforcement learning processes of the actor-critic capture the trial-and-error processes shaping the functioning of striato-cortical loops involving the frontal eye fields [18].

III. Results

The model has been tested in various conditions: the conditions and the results are now described in detail.

A. Model performance

The model performance was tested with only the presence of the causal light or with both lights. For each of these two conditions, the model was run two times for 8000 steps to evaluate its learning capabilities: the first time learning was blocked for the whole simulation, while the second time it was blocked until 4000 steps and then released. Fig. 6 shows the cumulated rewards in the resulting four conditions. The figure shows that when learning is released the capacity of the model to obtain a movement-based reward rapidly increases in about 1000 steps and achieves a steady state after about 2000 steps.

The model learns in relatively few saccades to look at the button from all other spheres. How does this fast learning take place? The bottom-up component, in particular the sensitivity for edges, leads the system to explore only highly informative portions of space, i.e. the spheres. The bottom-up orientation reflex triggered by the sudden appearance of the lights brings the system to immediately foveate them when they appear. Overall, the bottom-up process focuses the system on relevant portions of space thus producing a 'clever exploration' of it. The resulting visual movements are then incorporated or overridden by the top-down learning process based on the system needs (rewards). This results in a great enhancement of the efficiency of such learning process [22], [24]. We now analyse the system learned behaviour in detail. To this purpose, we focus on few more informative experiments.

B. Experiment 1: both lights, no learning

In this experiment, when the system looks at the button the causal light is switched on and the random light is switched on with a probability $p = 0.005$ at each step. The top-down

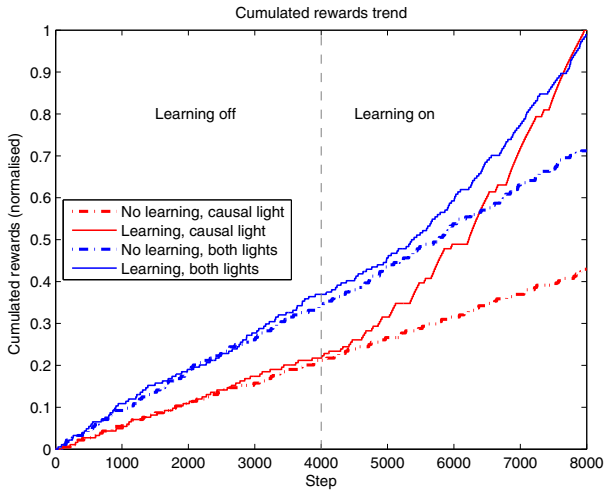


Fig. 6: Normalised cumulated rewards with only the causal light (red lines) and with both lights (blue lines) when learning was either blocked or blocked until 4000 steps and then released.

component is artificially disabled and so learning does not affects the system performance. As shown in Fig. 7, the system lingers on the six spheres with high probability and looks at the lights when they appear. Interestingly, a slight bias on the button position is already observed. This is due to the IOR: the causal light attracts the system attention immediately after it looks the button, so the system remains on such sphere for a shorter time in comparison to the other spheres and hence IOR charges less in correspondence to its position. This facilitates a saccade on the button after the causal light fades away.

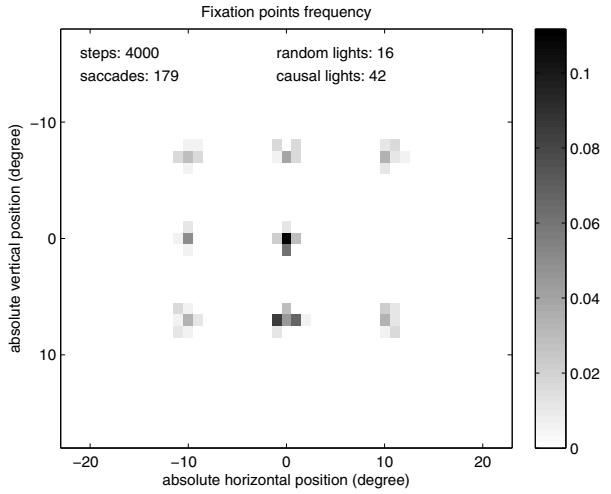


Fig. 7: Experiment 1: spatial frequency of saccades with the two lights and no learning. The figure also indicates the steps of the simulation, the number of saccades performed, and the random and causal lights switched on during the simulation.

C. Experiment 2: only random light, learning

In this experiment, only the random light is present and learning is on. As shown in Fig. 8, the top-down component cannot learn any meaningful skill as there is no statistical regularity linking the system behaviour and the random light appearance. Indeed, the little knowledge accidentally learned because of the appearance of the random light, and associated to one of the six spheres, is not supported by regular repetitions of the same experience as there is no stable action-effect contingency. As a consequence, the model unlearns that knowledge and so does not acquire any stable behavioural bias. The system thus remains controlled only by the bottom-up component and so continues to visit all the six spheres with a similar frequency, while temporarily looking at the random light when it appears on the basis of the movement reflex.

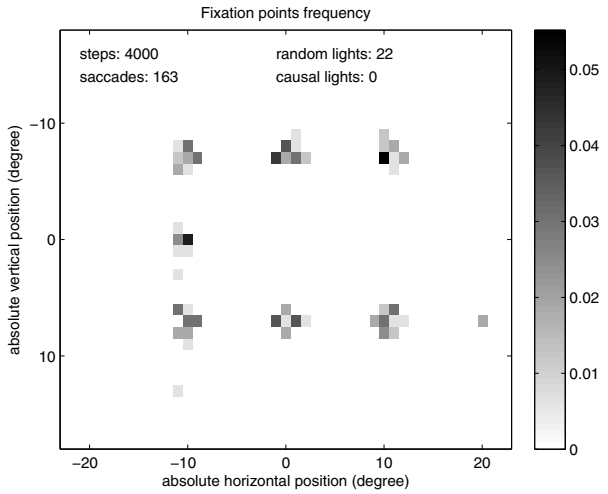


Fig. 8: Experiment 2: spatial frequency of saccades when only the random light is present and the system can learn.

D. Experiment 3: both lights, learning

In this experiment, both lights are present and learning is on. As shown in Fig. 9, after learning the system spends most time alternating saccades between two spheres: the button and the causal light. The top-down component indeed soon discovers and learns the skill to cause a change in the world, i.e. looking at the button for switching the causal light on. Due to noise and IOR the system sporadically looks at other spheres, but then immediately comes back to the button from them.

Interestingly, the presence of the random light does not deteriorate the system performance, see Fig. 9. Indeed, the appearance of the random light sometimes attracts the attention of the system, but this then returns to pay attention to the button and the causal light.

E. System functioning

Fig. 10 presents a snapshot of the activation of some key components of the model, and explains their role for selecting saccades target. This allows a better understanding of the model functioning underlying the behaviours described above.

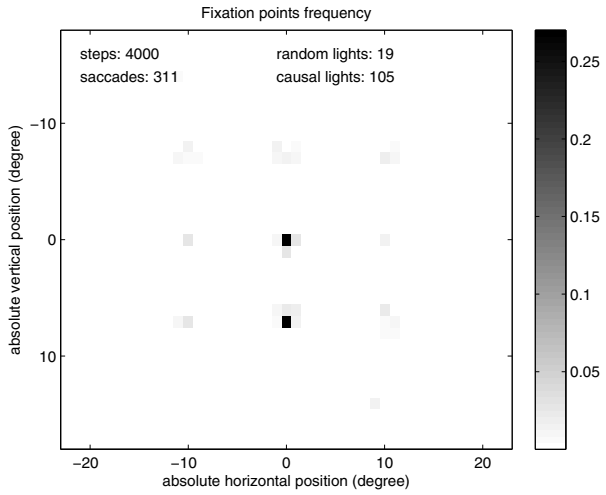


Fig. 9: Experiment 3: spatial frequency of saccades with both lights and learning.

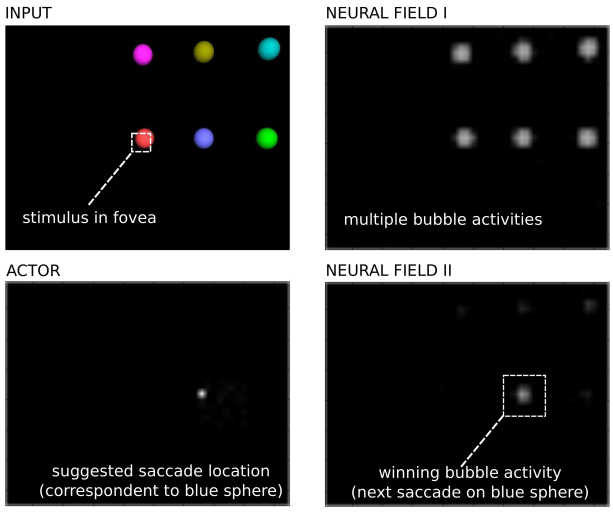


Fig. 10: Activation of some key maps of the model in an important phase of the environment exploration. From top-left in clockwise order: current input (the red sphere); activation of neural field I showing six bubble activities centred on the six spheres; activation of neural field II showing the formation of a winning bubble activity; activation of the actor showing the saccade target 'suggested' by the top-down component when the input is the red sphere, i.e. 'look to the right' (to reach the button).

IV. Conclusion

We presented a the first version of a new bio-inspired model, part of a research agenda directed to study the brain mechanisms underlying intrinsically motivated learning of attentional behaviours. The novelty of the model resides in its link with the brain attentional system of primates pivoting on the superior colliculus. The model captures the capacity of this system to control eye movements based on retinal and cortical inputs, thus integrating both bottom-up and top-down information. More importantly, the model also captures some aspects of superior colliculus contribution to generate intrinsically motivated learning signals on the basis of movement detection. The tests of the model showed the model soundness, in particular that it can learn attentional skills capable of actively causing changes in the environment while ignoring changes not depending on the agent's behaviour.

The tests of the model suggest several possible future enhancements of the model. These enhancements represent the next steps of the research agenda within which the model was developed, and that we now illustrate.

First, the model used a hardwired inhibition of return to foster exploration. This could be avoided by exploiting the *object-related inhibition of return emerging from using an actor capable of inhibiting*, and not only exciting, the possible target locations for saccades [22].

Second, currently the primary reward generated by movement detection never fades away, so the system repeats the learned visual skill forever. As mentioned in Section I, the *learning signal produced by the biological superior colliculus fades away when the system learns to predict it* thus leading the system to engage with different activities. Adding this function to the model would fully capture the functionality of prediction-based intrinsic motivations.

Third, the previous point raises the need to endow the model with the capacity to *save the learned skill* once engaging with another activity, and to recall such skill if in the future the related effect becomes desirable. This would imply to endow the model with the capacity to represent the changes of the world in cortex (goals), and to link them to the skills to accomplish them through inverse models [19], [20].

Fourth, it would be important to add to the model a *novelty-based intrinsic motivation system*, i.e. a system with the capacity of producing a reward signal on the basis of the novelty of the observed images/objects. Biologically, this would model the capacity of some areas of brain, such as the hippocampus, to respond to novelty [38]. This addition might be based on the SOM network component learning to recognise different inputs [39]. This addition would allow the study of the interplay between prediction-based and novelty-based intrinsic motivations [25].

Fifth, it would be important to *test the model with real cameras and images*. Pilot tests with a motorised web-come show that this is indeed possible if some of the additions above are performed, in particular the introduction of the emergent inhibition of return capable of disengaging the system from regions of the image that are highly attractive for the bottom-up component.

Last, here we used a task where eye movements directly cause changes in the environment. In the future we will integrate the visual model with further *components that control robotic arms* so as to study their interplay [22], [23].

ACKNOWLEDGMENT

This research has received funds from the European Commission under the 7th Framework Programme (FP7/2007-2013), ICT Challenge 2 "Cognitive Systems and Robotics", project "IM-CLeVeR - Intrinsically Motivated Cumulative Learning Versatile Robots", grant agreement no. ICT-IP-231722.

REFERENCES

[1] G. Baldassarre and M. Mirolli, Eds., *Intrinsically motivated learning in natural and artificial systems*. Berlin: Springer, 2013.

[2] G. Baldassarre, "What are intrinsic motivations? a biological perspective," in *Proceedings of the International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2011)*, 2011, pp. E1–8, Frankfurt am Main, Germany, 24-27 August.

[3] P.-Y. Oudeyer and F. Kaplan, "What is intrinsic motivation? a typology of computational approaches." *Frontiers in Neurorobotics*, vol. 1, no. 6, 2007.

[4] G. Baldassarre and M. Mirolli, "Intrinsically motivated learning systems: An overview," in *Intrinsically Motivated Learning in Natural and Artificial Systems*, G. Baldassarre and M. Mirolli, Eds. Berlin: Springer-Verlag, 2013, pp. 1–14.

[5] A. Barto, M. Mirolli, and G. Baldassarre, "Novelty or surprise?" *Frontiers in Psychology – Cognitive Science*, vol. 4, no. 907, pp. e1–15, 2013.

[6] J. Schmidhuber, "Curious model-building control systems," in *Proceedings of the International Joint Conference on Artificial Neural Networks*, vol. 2, 1991, pp. 1458–1463.

[7] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, 2007.

[8] J. Schmidhuber, "Formal theory of creativity, fun, and intrinsic motivation (1990-2010)," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 230 –247, 2010.

[9] A. G. Barto, S. Singh, and N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills," in *International Conference on Developmental Learning (ICDL2004)*, 2004, pp. 112–119, la Jolla, CA.

[10] M. Schembri, M. Mirolli, and G. Baldassarre, "Evolving childhoods length and learning parameters in an intrinsically motivated reinforcement learning robot," in *Proceedings of the Seventh International Conference on Epigenetic Robotics (EpiRob2007) Modeling Cognitive Development in Robotic Systems*, 2007, pp. 141–148, piscataway, NJ, 5-7 November 2007.

[11] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Which is the best intrinsic motivation signal for learning multiple skills?" *Frontiers in Neurorobotics*, vol. 7, pp. e1–22, 2013.

[12] A. Baranes and P. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 49–73, 2013.

[13] A. Pitti, H. Mori, S. Kouzuma, and Y. Kuniyoshi, "Contingency perception and agency measure in visuo-motor spiking neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 1, no. 1, pp. 86–97, 2009.

[14] E. Polizzi di Sorrentino, G. Sabbatini, V. Truppa, A. Bordonali, F. Taffoni, D. Formica, G. Baldassarre, M. Mirolli, E. Guglielmelli, and E. Visalberghi, "Exploration and learning in capuchin monkeys (sapajus spp.): the role of actionoutcome contingencies," *Animal Cognition*, pp. e1–8, 2014.

[15] F. Taffoni, E. Tamilia, V. Focaroli, D. Formica, L. Ricci, G. Di Pino, G. Baldassarre, M. Mirolli, E. Guglielmelli, and F. Keller, "Development of goal-directed action selection guided by intrinsic motivations: an experiment with children," *Experimental Brain Research*, vol. 232, no. 7, pp. 2167–2177, 2014.

[16] P. Redgrave and K. Gurney, "The short-latency dopamine signal: a role in discovering novel actions?" *Nature Reviews Neuroscience*, vol. 7, pp. 967–975, 2006.

[17] P. Redgrave, K. Gurney, and J. Reynolds, "What is reinforced by phasic dopamine signals?" *Brain Research Reviews*, vol. 58, no. 2, pp. 322–339, 2008.

[18] O. Hikosaka, Y. Takikawa, and R. Kawagoe, "Role of the basal ganglia in the control of purposive saccadic eye movements," *Physiol Rev*, vol. 80, no. 3, pp. 953–978, 2000.

[19] G. Baldassare, F. Mannella, V. G. Fiore, P. Redgrave, K. Gurney, and M. Mirolli, "Intrinsically motivated action-outcome learning and goal-based action recall: A system-level bio-constrained computational model." *Neural Networks*, vol. 41, pp. 168–187, 2013.

[20] G. Pezzulo, G. Baldassarre, M. V. Butz, C. Cristiano, and J. Hoffmann, "From actions to goals and vice-versa: theoretical analysis and models of the ideomotor principle and tote," in *Anticipatory Behavior in Adaptive Learning Systems: From Brains to Individual and Social Behavior*, M. V. Butz, O. Sigaud, G. Pezzulo, and G. Baldassarre, Eds. Berlin: Springer-Verlag, 2007, pp. 73–93.

[21] G. Baldassarre and M. Mirolli, "Deciding which skill to learn when: Temporal-difference competence-based intrinsic motivation (TD-CB-IM)," in *Intrinsically Motivated Learning in Natural and Artificial Systems*, G. Baldassarre and M. Mirolli, Eds. Berlin: Springer-Verlag, 2013, pp. 257–278.

[22] D. Ognibene and G. Baldassare, "Ecological active vision: four bio-inspired principles to integrate bottom-up and adaptive top-down attention tested with a simple camera-arm robot," *IEEE Transactions on Autonomous Mental Development*, In press.

[23] D. Ognibene, G. Pezzulo, and G. Baldassarre, "How can bottom-up information shape learning of top-down attention-control skills?" in *Proceedings of 9th IEEE International Conference on Development and Learning (ICDL2010)*, 2010, pp. 231–237, ann Arbor, MA, 18-21 August.

[24] R. Marraffa, V. Sperati, D. Caligiore, J. Triesch, and G. Baldassarre, "A bio-inspired attention model of anticipation in gaze-contingency experiments with infants," in *IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2012)*, 2012, pp. e1–8, San Diego, CA, 7-9 November.

[25] D. Caligiore, M. Mustile, D. Cipriani, P. Redgrave, J. Triesch, M. De Marsico, and G. Baldassarre, "Intrinsic motivations driving learning of eye movements: an experiment with human adults," *PLoS ONE*, subm.

[26] Q. Wang, J. Bolhuis, C. Rothkopf, T. Kolling, M. Knopf, and J. Triesch, "Infants in control: Rapid anticipation of action outcomes in a gaze-contingent paradigm," *PLoS ONE*, vol. 7, no. 2, 2012.

[27] R. Shadmehr and S. P. Wise, *The computational neurobiology of reaching and pointing: a foundation for motor learning*. Cambridge, MA: The MIT Press, 2005.

[28] E. Kandel, J. H. Schwartz, and T. Jessel, *Principles of Neural Science, 4th ed.* McGraw-Hill, New York, 2000.

[29] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[30] B. White and D. Munoz, "The superior colliculus," in *Oxford Handbook of Eye Movements*, S. Liversedge, I. Gilchrist, and S. Everling, Eds. University Press, 2011, pp. 195–213.

[31] R. Krauzlis, L. Lovejoy, and A. Znon, "Superior colliculus and visual spatial attention," *Annual Review of Neuroscience*, vol. 36, pp. 165–182, 2013.

[32] T. Anastasio, *Tutorial on Neural System Modeling*. Sunderland, MA: Sinauer, 2010.

[33] J. Lupianez, R. Klein, and P. Bartolomeo, "Inhibition of return: Twenty years after," *Cognitive Neuropsychology*, vol. 23, no. 7, pp. 1003–1014, 2006.

[34] R. Andersen, L. Snyder, D. Bradley, and J. Xing, "Multimodal representation of space in the posterior parietal cortex and its use in planning movements," *Annual review of neuroscience*, vol. 20, pp. 303–330, 1997.

[35] C. Galletti, P. Battaglini, and P. Fattori, "Parietal neurons encoding spatial locations in craniotopic coordinates," *Experimental brain research*, vol. 96, no. 2, pp. 221–229, 1993.

[36] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press, 1998.

[37] C. Pierrot-Deseilligny, D. Milea, and M. R.M., "Eye movement control by the cerebral cortex," *Current Opinion in Neurobiology*, vol. 17, pp. 17–25, 2004.

[38] J. E. Lisman and A. A. Grace, "The hippocampal-VTA loop: controlling the entry of information into long-term memory." *Neuron*, vol. 46, no. 5, pp. 703–713, 2005.

[39] S. Marsland, U. Nehmzow, and J. Shapiro, "A real-time novelty detector for a mobile robot," *arXiv preprint cs/0006006*, 2000.