Ecological Active Vision: Four Bioinspired Principles to Integrate Bottom–Up and Adaptive Top–Down Attention Tested With a Simple Camera-Arm Robot

Dimitri Ognibene and Gianluca Baldassare

Abstract—Vision gives primates a wealth of information useful to manipulate the environment, but at the same time it can easily overwhelm their computational resources. Active vision is a key solution found by nature to solve this problem: a limited fovea actively displaced in space to collect only relevant information. Here we highlight that in ecological conditions this solution encounters four problems: 1) the agent needs to learn where to look based on its goals; 2) manipulation causes learning feedback in areas of space possibly outside the attention focus; 3) good visual actions are needed to guide manipulation actions, but only these can generate learning feedback; and 4) a limited fovea causes aliasing problems. We then propose a computational architecture ("BITPIC") to overcome the four problems, integrating four bioinspired key ingredients: 1) reinforcement-learning fovea-based top–down attention; 2) a strong vision-manipulation coupling; 3) bottom–up periphery-based attention; and 4) a novel action-oriented memory. The system is tested with a simple simulated camera-arm robot solving a class of search-and-reach tasks involving color-blob "objects." The results show that the architecture solves the problems, and hence the tasks, very efficiently, and highlight how the architecture principles can contribute to a full exploitation of the advantages of active vision in ecological conditions.

Index Terms—Bottom-up top–down overt attention, camera-arm robot, ecological active vision, eye-hand coupling, inhibition of return, memory, partial observability, reinforcement learning.

I. INTRODUCTION

IN PRIMATES, vision is the richest source of sensory information, engaging a large part of the cortical brain machinery [48]. Visual information is perfectly suited to guide actions directed to accomplish goals, in particular to support

D. Ognibene is with the Center for Robotics Research, Department of Informatics, Kings College London, Strand, London WC2R 2LS, U.K. (e-mail: dimitri.ognibene@kcl.ac.uk).

G. Baldassarre is with Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Laboratory of Computational Embodied Neuroscience (CNR-ISTC-LOCEN), I-00185 Roma, Italy (e-mail: gianluca.baldassare@istc.cnr.it).

manipulation behaviors directed to change the environment. Vision can give information on the nature of resources in the environment, in particular about their utility for the animal's needs. Moreover, it can furnish a wealth of information on the size, orientation, location, and other features of such resources, and the spatial relations between them, needed to support the online control of action. In this respect, after the processing stages extracting basic visual information, the visual brain machinery splits in two major neural information pathways [42], [71]: (a) the ventral neural pathway, informing the animal on the nature of resources in the environment and hence supporting the trial-and-error acquisition of the capacity to select suitable actions; (b) the dorsal neural pathway, guiding the online control of action execution, e.g., to reach an object in a specific location.

Visual information richness, however, is double faced as the continuous visual flow can easily overwhelm the computational resources of the agent. This has been clearly shown by the difficulties encountered by classical approaches to vision that, following the initial Marr's proposal [67], attempt to use the whole available visual information to form a complete task-independent internal representation of the outer world [23].

Vision thus poses a difficult problem that we contribute to face in this paper from a perspective relevant for primates and humanoid robots: *how can an embodied agent exploit the overwhelmingly rich visual information from the environment to accomplish useful manipulations within it*? The approaches of *interactive/active vision*, inspired to what happens in primates, offer a general solution to this problem [3], [5], [12], [13], [23], [38]: *use a small fovea with a high resolution, and move it on relevant points in space to gather only the information needed to accomplish the task at hand*. Here we work on this solution from an "ecological perspective," hence we talk of *ecological active vision*. With this expression we mean that we develop the active vision solution following this strategy. First, we try to identify the specific problems encountered by the active vision solution when it is employed within "complete" embodied systems that do not use vision and attention only to collect information on the environment, but also to best guide their pragmatic interactions with it in order to change it and obtain useful resources, similarly to what primates do when acting in a "naturalistic" (versus abstract/engineered) conditions. Second, we seek solutions to those problems by looking at the solutions found by nature, both in terms of behavior and, when possible, brain anatomy and physiology (our solutions are thus "bioinspired," see below).

This paper starts to face the visual-richness problem illustrated above from the fact that fovea-based active vision can indeed solve the problem but in ecological conditions it also opens up four critical new problems:

1) Since the information relevant for the system depends on the pursued goal/faced task [90], [115], the system has to *learn where to look* on the basis of it. In particular, as in ecological conditions the environment is always changing, the system has to capture the *structure of the class of conditions* where it pursues the goal and generalize over their variable features [103]. How can this learning process be implemented and guided?

2) Detailed information is available at the location explored with the fovea, while the feedback (e.g., rewards) supporting eye-movement learning is accomplished by manipulation actions possibly performed in a different location: how can the system solve this *spatial indirectness problem* [95]?

3) The system has to learn where to look to support manipulation actions while only these actions can change the environment and generate learning feedback for learning to look. How can visual learning take off and solve this challenging *circular-dependency learning problem*?

4) The fovea, working together with a low-sensitivity periphery, gates away a large part of visual information and thus causes a difficult *aliasing problem* [113] for which different objects/conditions in the scene can perceptually appear the same: how can this problem be solved?

The literature has faced these problem largely in isolation from each other (see Section IV for a review). Instead, this paper proposes a system that *faces them in an integrated fashion*. The rationale for doing so is that facing these problems together leads to building a system that solves them in novel, more effective ways.

The system we propose is based on four general bioinspired ingredients. By "bioinspired'" and "general" we mean here that: (a) the ingredients are "bioinspired," rather than "bio-constrained," in the sense that they do not aim to capture detailed mechanisms of brain, or processes of behavior, as measured in specific empirical experiments: indeed, the ingredients were isolated by looking at how primates solve the four problems by abstracting, with a computational perspective, the knowledge from several specific papers on these issues (the contributions of some of these papers are specified in the following sections); (b) the ingredients aim to capture some "general" principles informing the organization of the attentional systems of primates: they should so be useful for different specific accounts of such systems. Based on these features, the principles are expected to facilitate a more detailed understanding of primate cognitive architectures and also to support the design of effective controllers for autonomous humanoid robots. The ingredients are as follows:

1) A top–down adaptive attention component capable of *learning by trial-and-error where to look* [10], depending on the already foveated objects, on the basis of the success of the system in accomplishing its goals [115], [90]. In autonomous agents, these goals are often not merely "visual goals," directed to gather information on the environment (e.g., to form representations or models of the world, cf.

[14]), but rather goals directed to *manipulate* and physically change the environment for own benefit. Here we focus on this latter class of goals, often overlooked by the literature on attention but very important for organisms as highlighted by the "attention for action" views [1], [10].

2) A *strong coupling between attention and manipulation actions*. The coupling manifests in relation to two key aspects of manipulation actions, namely their selection and their execution. In particular, in the system: (a) vision supports the *selection* of manipulation actions: these are triggered when a location is looked for long, similarly to what happens consistently in human babies [29]; (b) the gaze direction furnishes some *parameters to manipulation actions* [2], [10], in particular on their critical "where" aspect [109]. The importance of the attention-manipulation coupling is also stressed and studied in another model developed with an approach that shares several features with ours [49].

3) A *bottom–up attention component* guiding attention on the basis of task-unrelated, but readily available, bottom–up information [54], [85]. This supports initial exploration when the system has not yet acquired top–down attention skills, and hence facilitate their learning. Moreover, after learning bottom–up attention furnishes the system relevant information on the specific variable aspects of the environment that cannot be captured by top–down attention.

4) A *novel memory system storing information on potentially useful visual actions* rather than on percepts [31], [50], [63]. Memory is a major means to face the aliasing problem (e.g., [87]): the action-based nature of the memory system used here produces additional advantages for *performing* visual actions and also for *learning* them. The component is based on a neural map and hence is called Potential Action Map (PAM).

Due to these components, the system is called "Bottom-up attention, Top-down attention (BITPIC), PAM, Coupling system," where the *I*s in the acronym recall the importance of the Integration of the Ingredients.

BITPIC is here tested within a simulated robotic setup. The setup is based on very simple visual stimuli, basically color blobs, and a simple motor plant avoiding motor redundancy problems. These choices were motivated by the need to keep the components of the system as simple as possible so as to simplify their analysis and keep the focus of the research on the overall interplay between the system components and the principles on which they are based. Although simple, the chosen setup is sufficient to support the main claims of the paper. The principles incorporated in the system are expected to scale up to more complex setups as they work at a higher level with respect to the specific implementations of the sensory and motor elements of the system components (Section V indicates specific ways to do this).

Previous works from the authors presented initial analyses of the four ingredients of BITPIC but did not offer the broad integrated view presented here. Thus, the work presented in [80] investigated the role of a bottom–up attention component for the adaptation of a top–down component (see also [68]), the one of [77] focused on the advantages of the action-based memory component for visual search, the one of [81] investigated how

an active-vision system can readapt to changing tasks, and the one of [78] focussed on the effects of using manipulation actions to gather information from the world (i.e., as *epistemic actions* [56]). The novelty of this paper with respect to these works is as follows: (a) the clear identification of the four problems illustrated above, stemming from the active vision solution, and their systematic use to analyze the system behavior and learning processes; (b) the clear identification of the four architectural principles and the systematic study of how they contribute to solve the four problems; (c) a deeper understanding, supported by new data analyses, of how the *interplay* of the four principles leads to solve the four problems and also generates some interesting emergent properties of the system (for example, the developmental history of the system).

The main results of the paper can be outlined as follows. First, they show that the main tenet of active vision, namely that a small and actively controlled fovea improves learning and performance, holds for the type of system and setup used here. The result on the improvement of learning processes is novel as active vision has been shown to enhance performance by reducing the needed computations, but its importance *to enhance learning*, in our case of visual action selection, has not been studied with experiments but only highlighted theoretically [4], [22], [75]. Then the results show in detail how the integrated interplay of the four ingredients of the architecture can indeed solve or ameliorate the four active-vision problems and allow a full exploitation of active vision advantages for learning and performance. In particular the results show: 1) The *trial-and-error learning process* of the top–down attention component lead to solve the task with an unexpected high learning speed. This thanks to the development of strongly *embodied representations* [27], [70], [73] that are: (a) *action dependent*, in particular incorporating the effect of the sensorimotor loops engaged with the environment; (b) *task dependent*; (c) *parsimonious*, capturing only information relevant to guide visual and manipulation actions to solve the task. 2) The *coupling between visual and manipulation actions solves the indirectness problem* very effectively as learning feedback (reward) can be associated with percepts and actions related to a small portion of space. 3) During the initial phases of learning, *goal-independent bottom–up attention can support visual exploration in important ways*. In this respect, one of the most interesting results of the research shows how the system undergoes a developmental trajectory in which the visual routines learned by the top–down visual adaptive control progressively and gracefully incorporate, or override, the bottom–up attentional biases when, respectively, useful or detrimental for the pursued goal (by "visual routines" we mean sequences of visual actions whose component actions are strongly related with the invariances of the environment, the task, and between them, see [108]; by "incorporate" we mean that the learned visual routines rely upon bottom–up biases). Bottom-up attention is also useful after learning to exploit information changing from trial to trial. 4) The *novel action-based memory system substantially limits the negative effects of the aliasing problems,* caused by the reduced scope of the fovea, based on (a) an *emergent* inhibition of return [57] and (b) a continuously updated map of the promising regions of space where to look. The PAM also contrasts the effects of
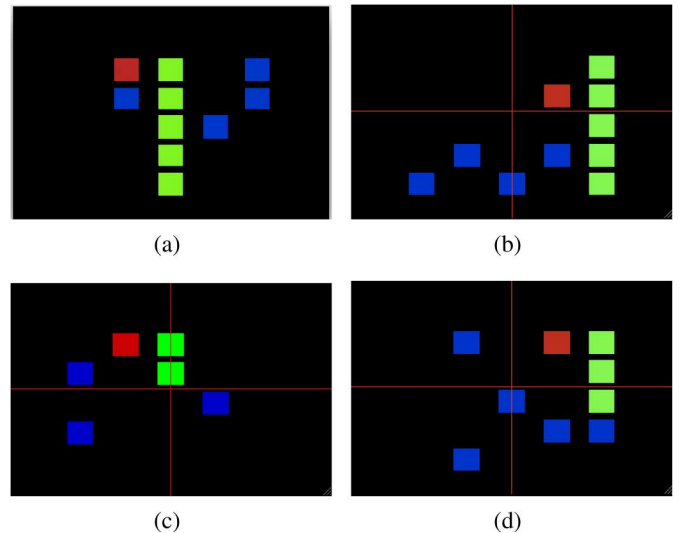

Fig. 1. (a-d) Examples of environments used in the experiment. Reprinted with permission from [80], Copyright IEEE (2010).

aliasing on learning, in particular by transferring visual action knowledge from object to object.

The rest of the paper is organized as follows. Section II illustrates the type of tasks and robotic setup used to test BITPIC, and gives the details on the system architecture and functioning sufficient to understand the results (the computational details for the reproduction of the architecture are presented in Appendix). Section III illustrates the results of the tests of BITPIC. Section IV discusses the results with respect to the problems illustrated above and the current literature. Section V draws the conclusions and illustrates possible future enhancements of the model.

II. THE TASK, ROBOTIC SETUP, AND ARCHITECTURE

A. The Simulated Robotic Setup

The setup used to test BITPIC is based on a kinematic simulation of the robotic setup illustrated in Fig. 2(a). This setup was first used in [76]. The setup is formed by a 4 degrees of freedom arm, acting on a 2D working plane formed by a computer screen, and a *fixed* camera looking down on the screen. The image acquired by the camera, overlapping with the computer screen working area, is used to form a larger image with a black background. A moving image centred on the system gaze is extracted from this larger image to simulate the moving-eye image which forms the input of the system. In this paper we used the simulated version of the system so as to run several experiments and for the theoretical scope of the paper, but the results are transferable to the real version of the setup, as shown in [76]. Indeed, the system is rather robust to different noise sources involved by the use of real robots and cameras as further explained in Section V (see also [75]).

B. The Task

The task used to test BITPIC is divided in trials each one ending when a reaching action is executed. In each trial the system "interacts" with a different "environment" *randomly*

drawn from class of possible environments, i.e. the images generated by the computer screen forming the arm working plane (Fig. 1). The variability of the images at each trial is important to test the system capacity to capture the *structure* of the class of environments, in particular the invariances of the spatial relations between objects. The images projected by the screen are built on the basis of a 6×5 (invisible) grid with "objects" (colored squares) located on some of its vertexes. The number and location of objects varies at each trial during which these objects are shown: 2 to 5 green cues having 100% luminosity and forming a randomly placed vertical line; 1 red target having 80% luminosity and placed at the left of the cue line; 2 to 5 blue distractors having 80% luminosity and randomly placed in the remaining grid vertexes; the distractors are not spatially related to the target. This structure of the task captures the need to visually explore the environment in search of resources (target) with which to interact with the arm, the possibility to exploit the eventual spatial relations existing between objects in the environment to guide visual exploration (cues), and the presence of non relevant objects (distractors).

At the start of each trial the system's gaze is randomly positioned in the working space. The goal of the system is to "touch" (i.e., be over) the red target with the tip of the arm ("hand") on the basis of the visual exploration of the environment. When the system performs a reaching action, the trial ends and the system receives a reward of 1 if it touches the target, and a cost of -0.1 if it reaches any point different from the target (another object or the empty space). A small reward of -0.005 is also given for each saccade at each simulation cycle. This structure of the reward captures conditions where the positive feedback from the environment follows only manipulation actions that change the world (e.g., procure food), while saccades are instrumental for them. In this way, eye movements receive positive feedback only indirectly on the basis of how they provide the arm with information relevant for solving the task at hand.

C. Overview of the System Architecture and Functioning

The architecture of BITPIC, shown in Fig. 2(b), integrates two neural controllers, one controlling the moving camera (simulated "eye"), and one controlling the arm.

The camera image is used to generate a high-resolution RGB small-sized *fovea image* and a low-resolution black-and-white large-sized *periphery image*. All components of the moving camera controller use a relative reference frame centred on the gaze direction, whereas the arm controller uses body-centred reference frames. The attention controller selects the next point to foveate by integrating information produced by three components. A first *bottom–up attention component* forms a hard-wired bottom–up *saliency map* [54], [92] resembling the function implemented by early stages of visual processing [85], [89]. This component receives as input the peripheral image and returns as output task-independent interesting locations where to look on the basis of their local information (here for simplicity we consider only luminance, but bottom–up saliency might also capture contrasts, movement, etc., [85]).

A second *adaptive top–down attention component* is based on an actor-critic reinforcement learning architecture [99] reminiscent of real brain basal ganglia [47] (trial-and-error processes implemented by basal ganglia play a critical role in the acquisition of voluntary eye movements, [46]). This component receives as input the fovea image, which informs the system on the foveated object, and returns as output multiple interesting locations where to look on the basis of a 2D "vote map'. Critically, this component learns to "vote" where to look on the basis of the task rewards caused by the arm actions, and so might for example learn to move away from distractors, exploit cues to find the target, keep fixating the target, etc. (note that, although the reward is sparse in time, learning takes place after each saccade/cycle).

The third component is a novel *Potential Action Memory (PAM)* based on a dynamic neural field. This component integrates in time the information ("votes") received from the top–down vote map and on this basis forms a map of locations in space where possibly look to solve the task. This function is reminiscent of the working memory processes involving the frontal eye fields (an area of prefrontal cortex) underlying the voluntary selection of saccade targets [31]. When the eye moves, the PAM content is shifted in the opposite direction to preserve its coherence with the new incoming votes, analogously to what happens in frontal eye fields [72]. Notice how the system does not need to memorize the bottom–up information as it is always readily available through the peripheral vision, nor to shift it as this is always anchored to the environment.

The output of the bottom–up component and the PAM component are integrated (they are summed up) in an final overall *saliency map*. The activation of this map is reminiscent of the activation of area LIP in parietal cortex where stimuli get only weakly activated (possibly on the basis of their bottom–up saliency) unless they are behaviorally salient for the task [43]. Based on this activation, at each step of the simulation the overall saliency map is first added some exploratory noise and then is used to select the next location where to foveate based on a neural competition (or, as here, a simplified winner-takes-all competition). This might result in a gaze shift or a fixation (reminiscent of microsaccades) [39], [61].

At the beginning of learning, the top–down component has very low votes and so saccades are mainly guided by the bottom–up component. However, with learning the votes of the top–down component can increase and overwhelm the activation of the bottom–up component. In this respect, if needed the top–down component outputs votes, which can also be negative, that can inhibit the effects of the bottom–up saliency of particular objects if these are not relevant for the task. The PAM allows to integrate (with a fading memory) top–down votes in time and so to integrate information on where to look on the basis of objects foveated at different times. For example, if a first object is foveated and suggests to look to a certain location L1 but the system looks at L2, the object foveated at L2 might strengthen the tendency to look at L1 and so this location might be eventually selected for the next saccade.

The arm controller was not the main focus of this research and so was trained before the main experiment. The arm controller selects the target for the reaching action through a neural competition. The reaching target is then fed to the position-control servos of the robot arm to produce the torques that lead the arm to reach the desired posture. This is similar to the muscles that drive limbs to reach desired *equilibrium points* [37]. The neural
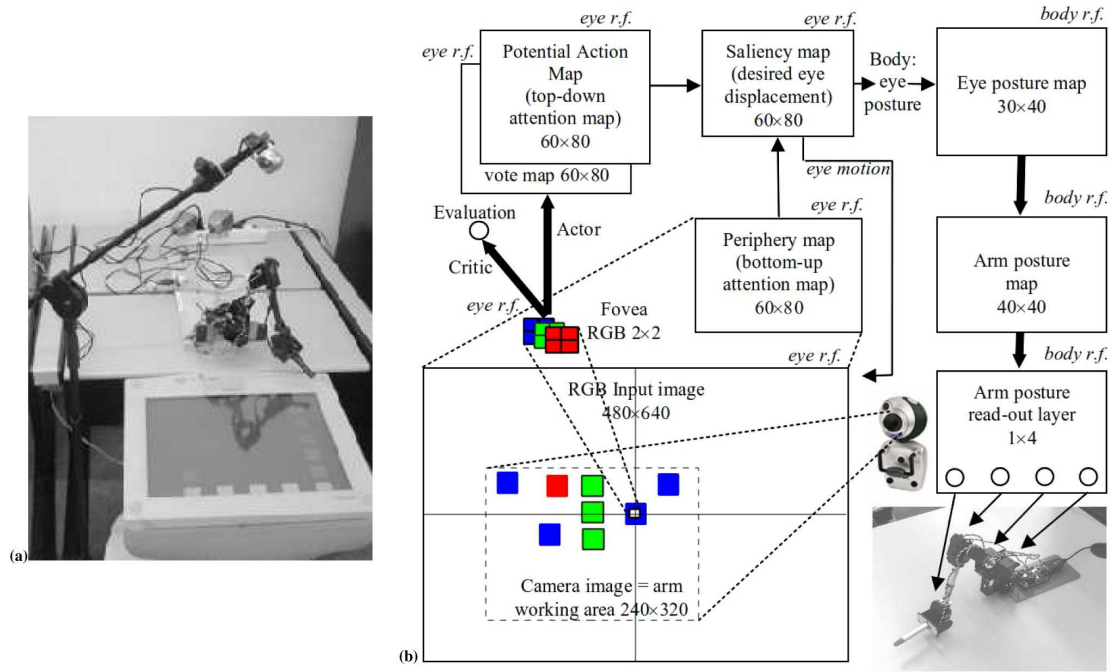
Fig. 2. (a) The robotic setup reproduced by the simulator and used to test BITPIC. The robotic setup is formed by an RGB web camera looking down on the working plane, formed by a CRT (Cathode Ray Tube) monitor which displays the stimuli of the tests, and a robotic arm formed by 3 segments and having 4 degrees of freedom. (b) The architecture of the system: the components on the left form the attention controller whereas the three components on the right (above the picture of the arm) form the arm controller. Adapted with permission from [76], Copyright by Springer (2008).

TABLE I
GROUPS OF SIMULATION TESTS OF THE MODEL, WITH INDICATION OF THE PROBLEM FACED, THE MAIN TESTED MODEL COMPONENT, THE MAIN RESULTS, THE DIFFERENT PARAMETERS USED (COMMON PARAMETERS ARE REPORTED IN TABLE III), AND THE TYPE OF MEMORY OF THE SYSTEM

Section	Problem faced	Key ingredient	Main results	Fovea size	Training trials no.	Memory type
Section III-A	0. Is small fovea advantageous here?	Small fovea	Learning enhancement	2, 4, 8, 16, 32, 48, 64, 80, 96	until ≤ 7 saccades	PAM
Section III-B	1. Need to learn where to look	Top-down attention	Fast learning; embodied representations	4	100	PAM
Section III-C	2. Spatial mismatch of reaching targets and attention focus	Attention-manipulation coupling	Manipulation effects can drive attention learning effectively	4	100	PAM
Section III-D	3. Looking-acting learning dependencies	Bottom-up attention	Bottom-up generates structured exploration for top-down learning; developmental trajectory	4	1000	PAM
Section III-E	4. Aliasing	Potential Action Map	Emergent inhibition of return; dynamic map of where to look	4	500	BASE, NEGPAM, PAM

map implementing the arm action competition is fed by the currently foveated location and thus the reaching movement to that location is triggered if the activation of a reaching target it encodes reaches a certain threshold, similarly to what happens in motor regions of brain during selection of reaching movements [26]. Note that in these conditions the eye-arm coupling is quite strong: reaching can be triggered only by a prolonged fixation and towards the fixated location. This was done to study the effects of the coupling without confounds, but in the future other sources of information might be introduced to activate the arm action competition map and trigger reaching to locations where the eye is not looking.

The eye-arm coupling of the model has two important features. First, it creates a strong link between the gaze and the reaching target as the arm can reach only targets that are foveated [1]. Second, it implies that the reaching movement is

triggered only if a particular object is foveated for a minimum period of time dependent of the threshold (e.g., 3-4 steps in a row in the current system), similarly to what happens in babies learning to reach [29], [30].

The arm controller also performs the visuo-motor transformations (inverse kinematic) needed to map the gazed location to the posture (angles) of the arm that causes the "hand" to be on the foveated location, similar to the sensorimotor transformation performed by the brain *dorsal visual pathway* at the level of the parietal cortex [21], [42].

Overall, during the task solution the whole system tends to exhibit the following behavior. At the beginning of a task, the bottom–up attention component leads the system to visually explore the various objects based on their visual objective saliency, for example to foveate the highly salient green objects. Once in a while the system fixates an object for long enough to trigger a

reaching movement to it and this produces a reward if the target is touched, or a small cost if it is not touched. The rewards and punishments collected by the system drive the reinforcement learning processes of the top–down attention component so that it learns to search useful information in the scene, for example to foveate the target for multiple steps so that a reaching movement is triggered towards it, or to exploit the sight of cues to move the eye to the target. These behaviors are described in detail in the sections below by analysing the results of the tests of the system.

The arm and eye controllers are based on bioinspired computational building blocks: (a) population codes, here implemented on the basis of 2D neural maps, used to represent sensory or motor information [86]; (b) dynamic neural-field networks, again implemented with 2D neural maps, used to integrate information in time and to perform winner-take-all neural competitions [36] (similarly to what happens in real brain cortex [25]); (c) bioinspired trial-and-error learning mechanisms based on the rare feedback from the environment (actor-critic reinforcement learning architecture, [99]).

III. RESULTS

This section presents the results of the tests of the model. The tests and results are organized in five groups summarized in Table I and described in the following five subsections.

A. The Role of the Fovea Size in Adaptive Active-Vision Systems

A first test was directed to investigate the effects on learning of different fovea sizes. A small fovea increases perceptual aliasing as the system looses the overall view of the whole scene. On the other side, it reduces the computational resources needed to process the image and drastically simplifies the input patterns to which the system has to learn to associate saccadic actions. What is the final balance of these contrasting costs and benefits for a system based on the four ingredients considered here? The relevance of this experiment resides in the fact that, as mentioned in Section I, the advantages of a small fovea size *for learning* have not been studied in depth in the literature.

To answer this question, the system was tested with different sizes of the fovea side: 2, 4, 8, 16, 32, 48, 64, 80, and 96 pixels. To have an idea on the actual scope of these sizes, consider that with a 96×96 pixel fovea the system can perceive at the same time 3×3 objects located in contiguous vertexes of the environment grid. For each fovea size, we ran five repetitions of the simulation. Each simulation lasted until the system learned a stable and effective visual strategy, corresponding to less or equal 7 saccades to finish the trial in 10 consecutive trials on average (the reason of the number 7 is that an optimal strategy requires on average about 3 saccades to find the target and 2-4 fixed saccades to trigger a reaching action). After this, we tested the performance of the system with learning disabled for 100 trials.

Fig. 3 shows the performance of the system with different fovea sizes after it has solved the task. The figure shows that: (a) the average reward that the system obtains in 100 trials after learning is high (up to 0.89) for small fovea sizes, and low for large fovea sizes; (b) the total number of trials needed to learn
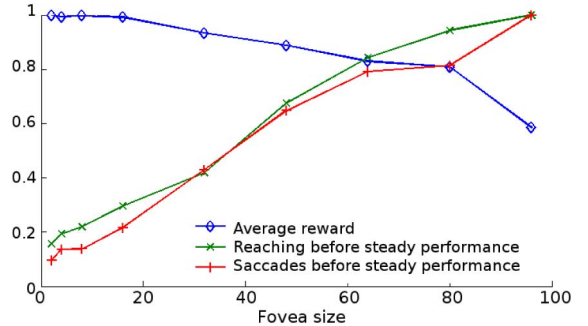

Fig. 3. Effects of different fovea sizes on the learning speed and performance of the system. The *x* axis reports the size of the fovea in pixels (2, 4, 8, 16, 32, 48, 64, 80, 96). The *y* axis reports three measures of performance and learning, normalized to 1, of the system having different fovea sizes. "Average reward": average reward over 50 trials. "Saccades before steady performance": number of saccades the learning process takes to find an efficient visual strategy (i.e., ≤ 7 saccades on average in the last 10 trials); "Reaching before steady performance": the total number of reaching actions necessary to acquire an efficient visual strategy. The three measures were normalized to 1 with their maximum levels being, respectively: 0.89 units of reward (obtained with a 4 pixel fovea size); 350 trials (obtained with a 96 pixel fovea size); 4010 saccades (obtained with a 96 pixel fovea size). Data refer to an average of 5 simulation runs performed for each different fovea size.

the task increases with the fovea size (up to 350); (c) similarly, the total number of saccades necessary to learn the task increases with the fovea size (up to 4010).

The key explanation of the efficient learning of the system with a small fovea size is that a small fovea avoids the need to learn to associate visual actions to images involving a large number of different possible combinations of objects and spatial locations. The result (a) implies that, even after learning to visually find the target, the system with large fovea sizes still triggers some reaching actions that miss the target. Direct observation of the system behavior shows the reason of this. Due to the residual partial observability of the scene and the high complexity of the image, the system with large fovea sizes sometimes falls in local minima that lead it to fixate nontarget objects and still trigger reaching. A possible improvement of this could probably be obtained with a higher exploration noise, but this might also have negative effects on the learning process and the steady performance.

1) Interim Discussion: The results on the effect of different fovea sizes are discussed here as they do not relate to the main issue of the paper, namely the four active vision problems and their solution, discussed in Section IV.

When the fovea size increases, the possible number of images increases exponentially with the number of pixels. However, Fig. 3 indicates that the performance of the system does indeed decrease with a larger fovea size, but much less than exponentially. What are the reasons of this? Aside the trivial explanation for which information of pixels is highly redundant and so increases less fast than the number of possible images, another explanation is that the advantages of a smaller fovea size are compensated by two factors. The first factor is visual aliasing: this requires the system to learn a *whole range of possible saccade targets* rather than one single possible target, as it would happen with an image covering the whole scene and without aliasing. A second factor, associated with the first, is that a smaller fovea requires solutions based on *sequences of*
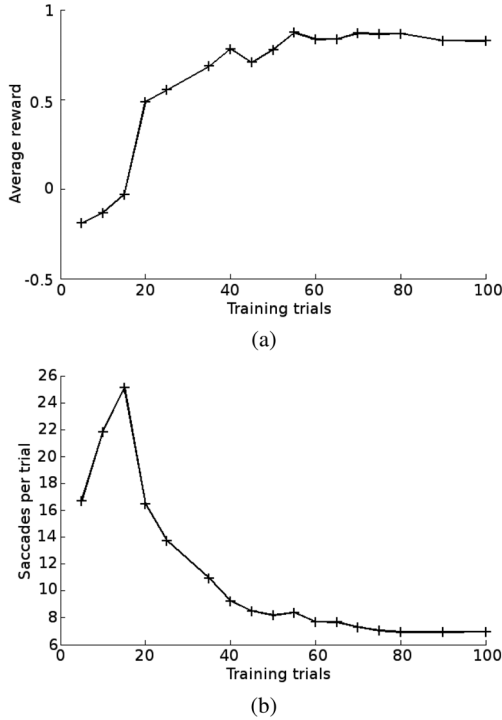
(a)

(b)

Fig. 4. (a) Evolution of average reward per training trial during the simulation. Note that if the system follows a random visual exploration, not guided by the bottom–up and top–down components, it gets a very low, negative reward due to the cost of saccades and as the chance to trigger a reaching to the target is very low (see next point "b"); the reward is instead about +0.34 if one assumes that the system immediately obtains the target reward as soon as it foveates the target, i.e. without the need to trigger the reaching action. (b) Evolution of average saccade number per trial. Note that with a random behavior the system takes about 108 saccades on average to foveate the target the first time (hence 108^4 to saccades to foveate the target 4 times in a row and trigger the arm reaching); with the bottom–up component it takes about 12 saccades to foveate the target the first time. For each marked number of trials, data in both "a" and "b" were collected by running the system for 1000 trials (during which learning was stopped) and then by averaging the resulting reward/number of saccades per trial. Such high number of trials was needed as the environment was different at each trial. Reprinted with permission from [80], Copyright IEEE (2010).
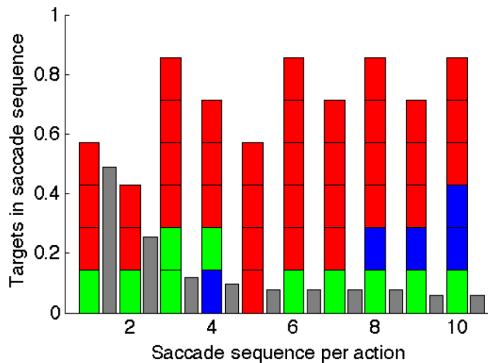

Fig. 5. Ten most frequent sequences of objects foveated during each trial. The green, red, and blue rectangles of each of the 10 columns indicate the objects (cue, target, and distractor, respectively) foveated during the trials. The sequences start form the bottom rectangle of each column and terminate with the top one (a target). On the right of each sequence-column, a grey bar indicates the frequency of the sequence (as percentage of all trials indicated by the y axis). The data refer to 1000 trials produced with a system previously trained for 100 trials and with blocked learning. If the initial eye position of a sequence was on the background, it is not reported in the sequences of the graph.

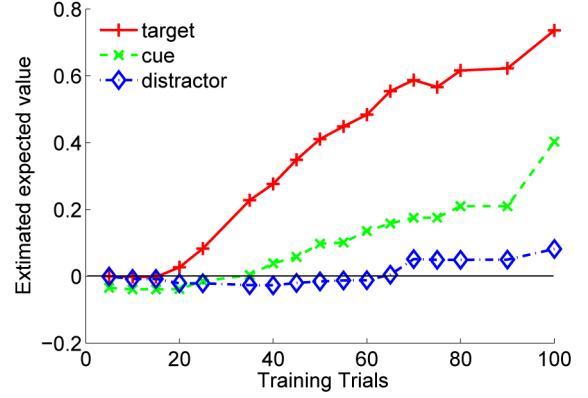*saccades*. Both aspects are further analyzed and discussed in the sections below.

Fig. 6. Evolution during learning of the critic's estimated expected value for the target, cues, and distractors.
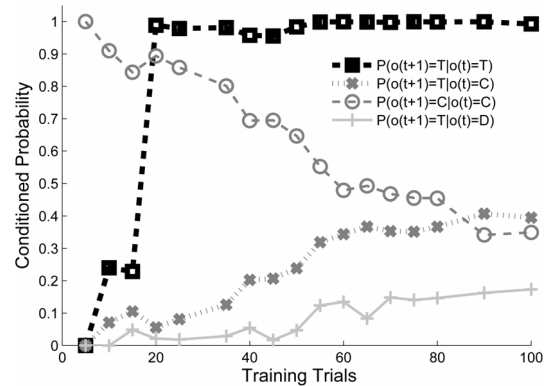

Fig. 7. Evolution during learning of the conditional probability of foveating a certain object after foveating another given object in the previous step. Symbols: T = target, C = cue, D = distractor, t = time index, $P(o)$ = probability of o. As an example, $P(o(t + 1) = T|o(t) = C)$ is the probability of observing T at time $t + 1$ given the observation of C at time t.

The small fovea leading to the best performance (2, 4 pixels) is rather small, about 0.001% of whole RGB input image. The larger human fovea (about 0.5-1% of the retina) is a compromise between the advantages granted by a small fovea, highlighted by the results reported above, and the need to have a larger fovea to recognize objects, as suggested by the fact that about half of visual cortex is dedicated to processing fovea information [116]. In the setup used here, objects can be unequivocally recognized on the basis of their color, so the object recognition problem can be solved with a small fovea. As in this condition a larger fovea does not bring any advantage, the advantages of its reduced size allows the system with a minimal one to outperform all others.

The results above suggest the possibility that the "fovea size" could change dynamically. Technically, this could be done through a component that gets the whole detailed input image as input and sends as output to the actor-critic a small "fovea image" having a *variable size*. Computationally, this would allow to tradeoff different advantages depending on the conditions faced, e.g., to exploit the advantages of a small fovea size for learning and the advantage of a large size for object recognition. Biologically, this would capture the adjustable size of the internal attention spotlight [35]. How the system could autonomously regulate, or learn to regulate, the fovea size is however an open issue.

These results, further corroborated by the analyses of the behavior and functioning of the system presented in the following sections, clearly show how a smaller fovea size can substantially enhance learning: to our knowledge, this is a novel result in the literature on active vision. Note that based on these results the fovea size was set to 2×2 pixels in all tests illustrated below.

B. Learning is Very Efficient and Acquired Visual Routines Capture the Structure of Object Spatial Relations

As illustrated in Section I, a first problem raised by the use of a small fovea is that it requires the capacity to learn to move the eye. In particular, this learning process should lead the system to acquire the *structure of the invariant spatial relations between objects which are relevant to solve the task at hand*. The test we now illustrate shows how the top–down learning component, integrated with the other ingredients, has indeed the capacity to acquire the needed visual routines to do that.

To this purpose, we trained the system for 100 trials. To monitor the system performance, every 5 trials of learning we got the system connection weights, froze the learning process, and measured the system performance for 1000 trials *each involving a different scene*. The data that we now illustrate are from the execution of only one run but the behavior of the agent was qualitatively similar in different runs.

Fig. 4(a) shows the average reward per reaching action received by the system during the 100 learning trials. Remarkably, *after only 20 trials* the performance reaches 0.5, and at about 60 trials it reaches the steady state (reward starts from a negative value as the system is punished when reaching fails to touch the target object). Fig. 4(b) shows how the evolution of the number of saccades per reaching action (trial). Initially the system takes about 16 saccades to accomplish the task. After 15 trials, the number rises to a maximum of 26 in correspondence to the maximum learning progress in reaching (Section III-E explains why). Then it progressively decreases to seven saccades per trial (the steady state). These results indicate that the development of the frequency of saccades and of the reaching actions is decoupled during learning even if the visual and reaching systems are closely coupled. In particular, the system exploits the cheap visual exploration to gain experience of the environment capitalizing on few expensive reaching actions.

Considering that, as shown below, the system takes three/four time steps (saccades) to trigger a reaching action, the results also indicate that after learning the system finds the target in about three saccades (seven in total) to manually reach the target. This high performance is accomplished by acquiring and using knowledge on the task-relevant spatial relations between objects and by exploiting the tight coupling between the eye and the hand. This is now explained more in detail.

Fig. 5 shows how, after learning, the system fovea manages to successfully explore the environment, and to support manipulation actions, on the basis of the object spatial relations and notwithstanding the environment variability at each trial. In particular, the figure shows the most frequent sequences of foveated objects exhibited in 1000 trials by a system previously trained for 100 trials. These sequences can be described and interpreted as follows:

Sequence 5. If the system is lucky and foveates a target, it immediately recognizes it and so fixates it until a reaching action is triggered to it.

Sequences 1-3 and 6-7. If the system first foveates a cue (this happens often due to its high luminosity), it then foveates an objects at the left of the cue and often this is the target: this behavior exploits the knowledge on the spatial relations existing between the cue and the target, acquired with learning by the top–down component, integrated with the hard-wired objective information of the bottom–up component on the location of (indistinct) objects along the left side of the cue line.

Sequence 4. If the system first foveates a distractor, it then quickly finds the cue by exploiting its visual salience and from there exploits the spatial relation of the cue with the target: note how in this case the high saliency of the cue is incorporated and exploited in the developed visual routine.

Sequences 8-10. After foveating the cue, the system might foveate a *distractor* rather than the *target*. Indeed, both might be at the correct position at the left of the cue, the fovea cannot see them, and the peripheral vision can only see the presence of objects but cannot discriminate their identity. If however the system does the error of foveating the distractor *it does not go back to the cue* but directly foveates another object at the left of the cue line and this is usually the target. This behavior is made possible by the memory of the PAM on the potential locations of the target stored when looking the cue. This shows the utility of the PAM component, in particular of its capacity to integrate information in time that allows the system to build up a map of potentially relevant locations so ameliorating the aliasing problem (this feature of the system is further considered in the following subsection).

These interpretations of the behavior of the system, in particular the fact that it relies on the knowledge on the invariant structure of the object spatial relations that the system acquired during learning, are corroborated by the analysis of Section III-D on the internal representations developed by the top–down component and their synergy with the bottom–up biases.

C. The Eye-Hand Coupling Solves the Indirectness Problem

In Section I we have discussed a second problem, the indirectness problem, consisting in the fact that in ecological conditions the fovea might not foveate the specific location where the arm causes the learning feedback. In this section we show that the coupling between the arm and the eye of the system solves this problem and plays a key role in the following development of the system.

Fig. 6 shows the evolution of the values assigned by the reinforcement learning process, namely by the critic of the top–down component, to the target, cue, and distractor. The figure shows how during learning the coupling between the arm target and the current gaze direction soon allows a direct transfer of the value of the reward obtained by the manipulation actions to the last foveated object before the reward, i.e., the target. This pivotal value transfer can then guide the whole remaining reinforcement learning process via the standard backward transfer of value from the target to other objects

in the environment seen before it, here the cues and even the distractors (for reasons explained below).

Fig. 7 furnishes further evidence on the importance of the eye-arm coupling to guide the whole developmental process and allow the system to learn the key spatial relations between the objects. In particular, the figure shows the evolution during learning of the conditional probabilities of foveating a certain type of object given a certain type of previously foveated object, for example, the probability that the system observes a target after observing a cue, $P(o(t+1) = T|o(t) = C)$, or a target after observing a distractor, $P(o(t+1) = T|o(t) = D)$. The figure focusses on the first critical 100 trials during which the behavior of the system rapidly improves (Fig. 4). The analysis of the figure indicates that:

- *Trial 1–15:* $P(o(t+1) = C|o(t) = C)$ *decreases.* Initially the system develops the skill of avoiding to stay on cues based on the negative rewards received when foveating them and attempting to reach them with the arm.
- *Trial 5–20:* $P(o(t+1) = T|o(t) = T)$ *abruptly increases.* The system quickly develops the skill of fixating the target: this is the critical phase in which the eye-arm coupling allows the system to rapidly transfer value from the manipulation realm to the visual target.
- *Trial 20–60:* $P(o(t+1) = T|o(t) = C)$ *substantially increases.* The system acquires the skill of moving from the cues to the target.
- *Trial 60–100:* $P(o(t+1) = T|o(t) = D)$ *takes off.* The system progressively learns to move from the distractors to the target (for the reasons explained below).

These results show that the eye-arm coupling allows the system to *rapidly and reliably transfer value from the realm of the manipulation actions*, where feedback derives in ecological conditions (imagine the reward is given by a food obtained with the arm) *to the pivotal object of the visual domain*, here the target foveal image. This step is the keystone of the reinforcement learning process giving rise to all the following developmental processes.

D. The Bottom-Up and Top-Down Attention Components are Highly Synergistic and Generate an Interesting Developmental Trajectory

Section I illustrated a third problem, a "circular-dependency learning problem", encountered by the active vision approach when used in ecological conditions: the system has to learn visual routines to support manipulation actions but only this actions can produce the feedback needed to guide the learning process of looking. This subsection illustrates how this problem is solved by a synergistic interplay between the bottom–up and top–down attention components, also supported by the coupling and the PAM memory. Consider that the problem of initially learning to look is hard as the possible inputs the system perceives are several, being them a combination of all possible locations in space and the possible objects. As we shall see, the bottom–up attention helps the bootstrapping of the learning process as it creates a strong initial focus on few locations and hence regularizes the initial experience.

The interaction of the bottom–up and top–down components solves the problem by generating an interesting developmental trajectory: exploration is initially guided by the "default"

bottom–up attention control, substituting the initial lack of top–down control, but then top–down control progressively shapes visual exploration to make it fully task dependent. After this process takes place, the visual routines exhibited by the system are organized on these principles: (a) the internal representations developed by the top–down control *capture the structure of the spatial relations* between objects by *generalising* over different environmental conditions: in detail, the top–down representations tend to drive saccades to *whole zones of space* where relevant objects might be; (b) the bottom–up biases allow the system to decide the actual locations where to look within such zones (thus, visual routines "incorporate" bottom–up biases that are useful to accomplish the task); (c) additionally, the nonuseful bottom–up biases are overridden by the top–down control with the aid of the PAM.

To show this in detail, we refer to a system that learns for 1000 trials in the same conditions illustrated in the previous subsections. Fig. 8 shows the activation of the vote map in critical phases of the learning process when the system foveates a cue, the target, or a distractor. The analysis of the graphs shows that the system generates the following developmental trajectory:

1) *Trial 0: Initial bottom–up exploration.* Initially the top–down component is "empty" and completely incapable of guiding visual exploration. In this critical phases, the bottom–up, task-independent component plays the function of guiding visual exploration to areas of the scene with a high visual salience, for example in areas of the space occupied by objects rather than by the homogeneous background.

2) *Trial 15: Inhibition of return of salient objects.* Under the guidance of the bottom–up saliency, the system explores with both the eye and the hand the objects with high visual salience, here the cues. The cost related to failed manual exploration generates a first interesting phenomenon: an *emergent inhibition of return* [57] that leads the eye away from cues when these are foveated (Fig. 8(a)). This process leads the top–down component to override nonuseful bottom–up biases, in particular: (a) to level the general attractiveness of the different objects in the environment independently of their bottom–up visual salience; (b) to progressively loose interest for objects that the system manipulates but that do not produce any relevant positive outcome for the task; this transition also strongly depends on the eye-arm coupling.

3) *Trial 20: Stay on target.* When visual exploration starts to involve the less salient objects, the system can visually and manually explore the target objects. In this case, the arm action produces a reward and so the system rapidly learns to fixate them when they are encountered (Fig. 8(e), (h); see Fig. 8(n) for a wholly formed map).

4) *Trial 40: Inhibition of return on less salient irrelevant objects.* With a more uniform exploration, the system also encounters less salient objects, such as the distractors, interacts with them, and eventually discovers their neutral nature. The system thus starts to develop an inhibition of return also for them [Fig. 8(i); see Fig. 8(o) for the wholly formed map].

5) *Trial 60: Learning of strong spatial relations between cues and target.* After enough trials, the system starts to discover

the spatial relations between the cues and the target, i.e., that the target is situated in a region at the left of the cue [Fig. 8(j); see Fig. 8(m) for the wholly formed map]. Notice how with learning the representation of the zone where the target might be located progressively enlarges, so allowing the system to generalize over different possible conditions. In each specific trial, the bottom–up component will then specify where to look within such zone (see below).

6) *Trial 1000: Learning of weak spatial relations between distractors and target.* After a prolonged learning, the system develops the tendency to look in two zones above or below the distractor when foveating it (Fig. 8(o); again, in each trial the bottom–up bias will specify where look within the two zones). The utility of doing this depends on the strong tendency of the system to look in the zone at the close left of the cue line, when foveating a cue, in search of the target (Fig. 8(j); see also Fig. 5): if while doing this the system foveates a distractor rather than the target, the tendency to look above or below the distractor will lead it to find the target with a high chance. This is an interesting case where the representation related to the distractor captures a spatial relation generated by the particular exploration strategy developed by the system, hence it is strongly action-dependent (see below for a further analysis).

We now show more in detail how the top–down component, aided by the PAM, has incorporated some bottom–up biases into the visual routines when they are useful for the task, or has overridden them when they are not useful. Fig. 9 shows the activation of the PAM during a typical sequence of three saccades targeting the cue, the distractor, and the target within the scene shown in Fig. 1(a). The figure shows that:

- *When on cue* [see Fig. 9(a)], the systems has acquired a strong tendency to go in a region at the left of the cue but the precise point within such region can be disambiguated only by the bottom–up component informing the system on the specific location of objects in such region: in so doing, the bottom–up component plays an important role to capture some features of the specific scene of the trial.
- *When on distractor* [see Fig. 9(b)], the systems has acquired a strong tendency to go up or down with respect to the distractor; notice how the PAM also sums up the tendency to go to the target from the previously foveated cue. Again this tendency covers two regions above and below the distractor, so the bottom–up component can disambiguate where to precisely look within them. At the same time, the strong bias of the bottom–up component to move to the highly salient cues at the right of the distractor is overridden by the PAM that encodes in memory the inhibition of return formed by looking the cue at the previous step [see Fig. 8(m)].
- *When on target* [see Fig. 9(c)], the system has acquired a strong tendency to remain on the target. At the same time, highly salient cues continue to be overridden by the PAM still storing in memory the inhibition of return for the cues.

E. The Potential Action Memory Contributes to Face the Aliasing Problems and to Support Learning

We have seen in Section I that a fourth problem introduced by the active vision approach is that the fovea gates away a lot of information and so creates a strong aliasing problem which might impair both the learning and functioning of the system. The aim of this section is to show how the PAM can give an important contribution to ameliorate this problem. To this purpose, we tested and compared three versions of the system:

1) *BASE system*. A system having all components of the system seen so far with the exception of the PAM. In this system the overall saliency map directly integrates the information from the actor vote map with the bottom–up information. The BASE system was not able to learn the task with a noise of 0.1 of the saliency map used in all experiments, so such noise was set to 0.5 (see Appendix for the details on the saliency map noise).

2) *NEGPAM system*. A system in which only the negative votes of the vote map are passed to the PAM (i.e., the PAM forgets the activation of positive units which are set to zero). This implies that the system can learn to avoid to foveate the same objects multiple times (inhibition of return, [57]) but it cannot accumulate positive activations in memory encoding the possible interesting regions of space to visit on the basis of the spatial relations between objects. The NEGPAM system was tested only with the original noise of 0.1.

3) *PAM system*. The complete BITPIC architecture used in the previous experiments, based on a fully functioning PAM component. The PAM system was tested with both 0.1 and 0.5 noise.

The environment used to test the three systems was the same as in the preceding sections. Two runs of 500 trials were performed for every condition. We now present the results in detail.

1) Analysis of Performance: Fig. 10 shows the success rate, measured as the proportion of trials ended with a successful reach to the target, of the three system versions during learning. Table II summarizes the performance and number of saccades of the three system versions after learning. After learning the three systems achieve a very similar performance: the proportion of trials ending with a successful reach over trials 450–500 is 0.96 for the BASE system, 0.97 for the NEGPAM system, and 0.98 for the PAM system. Similarly, the average number of saccades per trial over trials 450–500 is 6.63 for BASE system, 6.42 for the NEGPAM system, 6.44/6.77 for the PAM system. However, the three versions of the system exhibit a quite different learning dynamics. In particular, the number of trials taken to reach a steady performance (7 saccades to reach the target) and the average number of saccades in trials 1–250 are, respectively, 315 and 30 for the BASE system, 88 and 17 for the NEGPAM system, and 31/34 and 8/10 for the PAM system.

Why such a different learning speed? Let us first compare the BASE and NEGPAM systems. Initially, all systems are attracted and remain on the highly salient cues. This generates an activation of the vote map related to the cues that is (Fig. 11): (a) negative at its centre, because occasional arm reaches directed to the cue are penalized by the small cost; (b) positive in regions around the centre, as each time the system moves from a negatively charged cue to other (discounted) negatively charged cue it receives a positive "TD-error" (see Appendix). Such map configuration causes repeated cycles of fixations between cues, on one side, and distractors and target, on the other. However, the NEGPAM system can accumulate in the PAM the inhibition

collected by visiting more cues in sequence and so it can more easily overcome the attraction of their bottom–up saliency so as to visually and manually explore the targets and distractors. Instead, as soon as the BASE system moves from a cue to a noncue object it tends to go back to a cue as it does not store the inhibition of return related to the cues. Incidentally, note that these processes are the reason of the increase of the number of saccades per trial in the initial phase of learning of each system [see Fig. 4(b) for the PAM system]. A larger fovea could avoid this problem but would also substantially increase the number of input configurations to consider.

Comparing the learning speed of the NEGPAM and PAM systems, Table II shows that the maximum number of saccades performed in one trial by the three systems is rather different [this peak of saccades is achieved in the first phase of learning, Fig. 4(b)]. This number is 153 for the BASE system, 263 for the NEGPAM system, and 130/75 for the PAM system. This is due to the fact that the PAM system can more readily learn to fixate objects and so trigger reaching actions towards them, in particular to targets, while the NEGPAM system cannot. The reason can be seen by considering Fig. 12 showing the activation of the PAM and NEGPAM systems when they foveate a target in the initial phases of learning. Notice how the PAM system, differently from the NEGPAM system, has an activation of the PAM which is *positive at the centre*, so the system will tend to continue to foveate the target at the next time step. Where does this "positivity" come from? We have seen above that both the NEGPAM and PAM develop a map for highly salient objects that is negative at the centre and positive in the periphery, so promoting the exploration of novel objects with respect to the foveated, highly salient objects. The point is that *such positive bias is remembered by the PAM* of the PAM system once the system foveates other objects, and so it creates the central positive activation shown in Fig. 12(a). This allows the PAM system to keep foveating the target until the first reaching action on it is triggered. Instead, once the NEGPAM system foveates the new object it forgets the positive activation related to it and so jumps to other non-inhibited objects.

2) The PAM Generates an Effective Exploration of Objects: There is another way through which the PAM facilitates learning: by biasing exploration in such a way that the acquisition of the spatial relations between some objects facilitate the learning of spatial relations involving other objects. An example of this is related to the fact that, through the PAM, the acquisition of the spatial relation between the cue and the target can facilitate the learning of the spatial relation between the distractor (at the left of the cues) and the target. In this respect, Fig. 13 shows the vote map activation of the three systems when they foveate a distractor after learning. The figure shows that the PAM system has learned the spatial relation between the distractor and the target (namely to search up or down it, see Section III-D), while the other two systems have done so only partially. The acquisition by the PAM system of the distractor-target subtle spatial relation is largely due to a biasing exploration process that we now illustrate. Once the PAM system has started to acquire the information on the spatial relation between the cue and the target [see Fig. 13(e)], when it foveates the cue it has a strong bias to look in the area at the left of the cue line. If it foveates a distractor, the PAM will still

have in memory the bias to explore the locations within the area at the left of the cue line. If this bias, aided by the bottom–up information (see Section III-D), leads the system to foveate the target, then the spatial relation between the just-seen distractor and the target is crystallized in the connection weights related to the distractor votes [see Fig. 13(f)]. In this way, the cue-target spatial relation can greatly facilitate the acquisition of the distractor-target spatial relation. A second example of this transfer involves the cues: for similar reasons, learning to foveate in a certain region at the left of a foveated cue facilitates learning to foveate other regions at the left of a cue if a second cue is foveated by mistake as the PAM could drive the system to find the target with a third saccade.

The learning process of the NEGPAM system is instead less efficient. When it fixates the distractor, it does not have the PAM storing past positive biases towards the target zones based on the seen cue, and thus has a similar chance of moving to the target or to other distractors. For this reason it fails to fully learn the distractor-target spatial relation [see Fig. 13(d)].

The BASE system, without memory, looks at the cues, then possibly moves to the distractor, and once there it does not remember the inhibition of return with respect to the cues, so tends to look back to them (this also explains why the BASE system tends to learn a very strong inhibition of return for cues, see Fig. 11). As a consequence, it will completely fail to learn the distractor-target spatial relation [see Fig. 13(b)].

IV. DISCUSSION

In this section we discuss the key results illustrated in the previous sections and relate them to the relevant literature and existing systems. For the sake clarity, we consider the results in relation to the four key ingredients of BITPIC, but at the same time we highlight how such results strongly rely on their *integration*.

A. The Top-Down Attention Component

A first result of this work is that the reinforcement learning process of the top–down attention component, integrated with the other ingredients, allows the acquisition of the visual routines needed to accomplish the task at hand with an unexpected speed. A first reason of the efficiency of the learning process relies on the formation of strongly "embodied" representations [27], [70], [73] developed by the top–down component, namely the "votes" of its actor when BITPIC foveates the different objects. We now see this more in detail (below we consider the second reason, i.e. the initial role of the bottom–up component).

First, the representations developed by BITPIC are strongly *action dependent* and *deictic*. In particular, they incorporate the sensorimotor interactions the system engages with the environment. One interesting example of this is the representation of the votes associated to the perception of the distractor, which in the last learning phase becomes a predictor of the target position. This representation is strongly dependent on the bias of the system, acquired at the beginning of learning, to look at the left of the cues and so possibly incur in a distractor. This agrees with other works showing that autonomously learned internal representations and information processing are strongly action dependent [64], [66], [70], [73]. The representations formed by

the system are also "deictic" [12], [13], in particular they capture the spatial relations between objects with reference to the currently foveated point, i.e. to the condition of the agent (see also [66]). In its simplicity, BITPIC is hence a paradigmatic example of how behavior, via the environment, can have strong effects on the representation of objects [110].

The sensorimotor solutions found by BITPIC strongly rely on the bottom–up and top–down biases, and the memory of them stored in the PAM. This allows the system to generate visual action sequences strongly coupled with the environment ("visual routines", [108]) that can ameliorate the non-Markovian/aliasing problems generated by the setup. In this respect, evolutionary robotics [74], using genetic algorithms that search *whole* behavioral sensorimotor solutions, has shown the power of sensorimotor behaviors to solve aliasing problems. Evolutionary techniques [32], [33], [73], [113] are generally more robust to perceptual aliasing problem than reinforcement learning (RL) algorithms (but see [16]) and have also been used to learn to solve very complex attention task (e.g., driving a robot along a path with landmarks [100] or a car in a simulator [34]). The results presented here show that a RL attention system may nicely solve some aliasing problems if integrated with bottom–up attention and a memory device (below we further discuss the contribution of the PAM to this).

Second, the representations acquired by the top–down component are strongly *task dependent*. In particular, they are fully directed to support visual exploration and manipulations directed to solve the tasks generated by the system ecology. For example, the representation related to the target guides the system to fixate and trigger a reaching action to it, while the cue and the distractor representations support the search of the target. As discussed below, also the development of these representations during learning initially involves objects more closely related to positive reward and then objects less related to it (i.e.: from the target to the cues and distractors). These results agree with the insight, supported by various authors (e.g., [102]) since [115], for which task-based top–down control is pivotal for attention in naturalistic contexts. Based on this idea, trial-and-error learning has been recognized to be critical for attention guidance [44] and thus has been modelled with a reinforcement learning system as here [10], [11], [102]. Our study contributes to this research with novel insights on the effects on learning of the bottom–up and top–down attention interplay (see below).

Third and last, the representations developed by the top–down component are strongly *parsimonious*. In particular, although they mainly capture the spatial relations between objects, they are sufficient to successfully solve the task. The small fovea support generalization, in particular in relation to cues located at different absolute positions or with respect to the relative spatial relations of objects. The reason is that a small fovea abstracts information by gating away aspects not related to the foveated object, thus favoring generalization.

These three features of representations constitute an important departure from the view for which perception should aim to build a complete representation of the world [67]. In this respect, and consistently with other active vision views [13], [34], the results reported here indicate that, at least in some common conditions, simple representations as those discussed here can

suffice to support adaptive behavior if aided by the other mechanisms as those incorporated by BITPIC.

B. The Attention-Action Coupling

In Section I we have seen that autonomous agents controlling vision and manipulation have to face the "indirectness problem". The problem is generated by the fact that in ecological conditions relevant positive reward feedback is usually produced by modifications of the environment caused with manipulation, not just by looking around (an important exception to this are social contexts, not considered here). Thus, the reward accomplished by manipulation actions might happen to be associated with the currently perceived objects although these might be unrelated to the manipulation targets. The results illustrated here show that imposing a strong coupling between the focus of attention and the target of manipulation actions represents a very effective solution to this problem. Indeed, this coupling allows BITPIC to immediately and rapidly carve the keystone of its learning process, namely to identify the object to target with manipulation, and to attribute to it a high value. This step allows a fast and neat transfer of value information from the manipulation realm to the visual realm so that the standard back propagation of value from the target to previously-explored objects and actions, typical of trial-and-error learning, can take place.

Related to this, it is interesting that the literature is collecting important empirical evidence on the existence of strong coupling between eye fixations and reaching targets in infant early reaching [29], [30]. In particular, these studies show that infants' reaching actions are directed to locations that are attended for at least one second of accumulated "looking time." Such perceptual-motor coupling is more likely to occur in 9 months old infants and is not present from the onset of reaching but forms rapidly over a six week period following it. The results presented here contribute to explain the computational importance of such coupling for learning. Another developmental robotic model, presented in [49], assumes a close attention-manipulation coupling and shows its importance to support the development of fundamental associative knowledge supporting eye-arm coordination.

The coupling used here is also an instance of the important principle of "attention for action" [1], [2], [10], for which a main function of attention is the guidance of pragmatic actions (e.g., manipulation actions). In particular, the focus of attention not only aids the triggering of reaching actions when specific spots are attended for long (see above), but it also furnishes critical parameters to the performed action, in particular the location in space of the target for reaching [109]. This agrees with evidence on the performance of complex behavioral sequences of actions in adult natural behavior, showing that eye fixations consistently precede and guide the sequences of manipulation actions directed to different objects [44].

The capacity of autonomously identifying target objects based on the feedback from manipulation effects is a relevant feature of BITPIC, and very important for autonomous systems operating in ecological conditions. This differs from other models of attention often assuming that the target is known by the system (e.g., [27]) or that a relevant positive reward is given for simply looking at the target (e.g., [69]).

C. The Bottom-Up Component

We have seen in Section I that an autonomous system has to face the "circular-dependency learning problem," which is common in ecological conditions, of learning to look to suitably guide manipulation actions and, at the same time, guiding learning to look with the feedback caused by the manipulation actions. The results presented here indicate that a close interaction between the bottom–up attention component and the top–down adaptive component can solve this problem very effectively. In particular, during the initial phases of learning of a task, when the top–down component does not know what to do and visual exploration would be random, the task-independent bottom–up attention component can guide visual exploration on salient regions of space. This "default" guidance can greatly favor the acquisition of task-dependent top–down biases through a number of effects that we now discuss in detail.

First, the bottom–up component leads BITPIC to foveate high-contrast locations. In the setup used here, this leads the system to explore objects rather than the dull background. In general, bottom–up saliency favors the focussing on regions of space where relevant objects are located since "objects" usually create visual discontinuities with the background [53], [18]. Bottom-up attention captures this regularity common to most tasks independently of the specific goals they involve.

Second, the bottom–up attention, being based on the intrinsic saliency of objects, creates a *structured exploration* of them, especially when it gets progressively coupled with top–down inhibition of return (see below). As observed by other authors in relation to similar setups, the coupling of system with the environment allows "the structuring of input, and produces statistical regularities" that the system can exploit [64]. In this respect, the structured visual exploration initially caused by the bottom–up component supports learning much better than the random exploration that the top–down reinforcement learning system might produce. The reason of this is that it tends to cause *regular visual exploration sequences* on the basis of the salience of the image elements. These visual exploration sequences can then be evaluated by the reinforcement learning top–down component, and hence discarded or incorporated into the task-oriented visual routines depending on their utility (see below). Overall, this greatly facilitates the learning process, and is in stark contrast with the slow speed of learning based on random exploration typical of reinforcement learning systems. In this respect, in another work [68] we showed how a different system, again endowed with bottom–up and top–down attention components, can learn very fast (in 1-3 trials) to solve a different looking task as the bottom–up component kind of "supervises" the top–down learning process. In so doing, the model reproduces the surprising learning speed observed in babies engaged in the same task [112].

Third, when the bias given by bottom–up attention is useful for the task, such bias is readily incorporated in the visual routines. This process is similar to the processes involving innate reflexes studied in developmental models, where the reflexes play a key role in the initial learning phases and are later incorporated or overwhelmed by more sophisticated control [17]. An example of this is the tendency of BITPIC to foveate objects, which favors the focus on the target object (as soon as the focus

of the highly salient cues is overcome by the top–down inhibition of return). In this respect, the results show that a small change in the votes to stay on the target object can lead to an abrupt improvement of learning as it can succeed to keep the eye fixed on it by summing up with the bottom–up bias (see [75] for some experiments showing the robustness of the model to a wider span of possible saliency values of objects). Another critical example is the fact that, to be able to generalize over different conditions, the top–down biases cover whole zones where relevant objects might be (e.g.: "at the left of the cue line," "above and below the distractor"). We have seen how in this cases the bottom–up biases play a key role to specify the location where to look within those zones.

On the other side, and this represents a fourth relevant point, when a bottom–up bias is not useful or even detrimental to solve the task at hand, the top–down control can overcome it, for example giving rise to an emergent top–down inhibition of return [57]. In this respect, we have see that an important phase of BITPIC development is the acquisition of the capacity to move away from highly salient objects, here the cues.

We think these effects of the bottom–up biases on the learning process of the top–down component, robustly observed in very different setups, are some of the most important results of this and previous works.

Overall, these results give also a relevant contribution, from the particular perspective of learning, to the debate on the importance of bottom–up and top–down processes in attention (but see also the information maximization approach that negates this dichotomy, [19]). On one side, this debate involves positions that, also based on psychological evidence (e.g., in relation to pop-out effects, [107]), propose models that assign a prominent role to bottom–up features guiding visual exploration on the basis of stimulus-based saliency maps [53], [58], [88]. On the other side, the debate sees positions stressing the pivotal role of top–down task-dependent processes to explain human visual attention [41], [65], [79], [90], [96], [102]. In line with empirical research that shows that bottom–up and top down processes are both important and give independent contributions to visual exploration [60], this work contributes to the debate by highlighting the importance of their interplay *for learning* by pinpointing some specific aspects of such interplay.

One of the most interesting results of the research shows how the aforementioned processes give rise to a developmental trajectory characterized by the presence of several phases. Thus, during learning BITPIC initially focuses on stimuli with high visual salience ("cues") and then develops inhibition for them; then learns to focus on the target of manipulation actions; this opens up the possibility of learning the regularities related to stimuli having strong spatial relations with the target ("cues"); finally, the emerged structured coupling of behavior with the environment allows the system to capture more subtle regularities in the environment (e.g., here in relation to distractors).

The staged development exhibited by BITPIC depends on the particular visual feature of stimuli, the regularities of the environment, and the architecture of the system. The development of the system is so caused by the interactions of multiple subsystems rather than by a single cause and manifests as a soft assembly of several changes enabling better adaptiveness. In this respect, the developmental trajectory followed by BITPIC
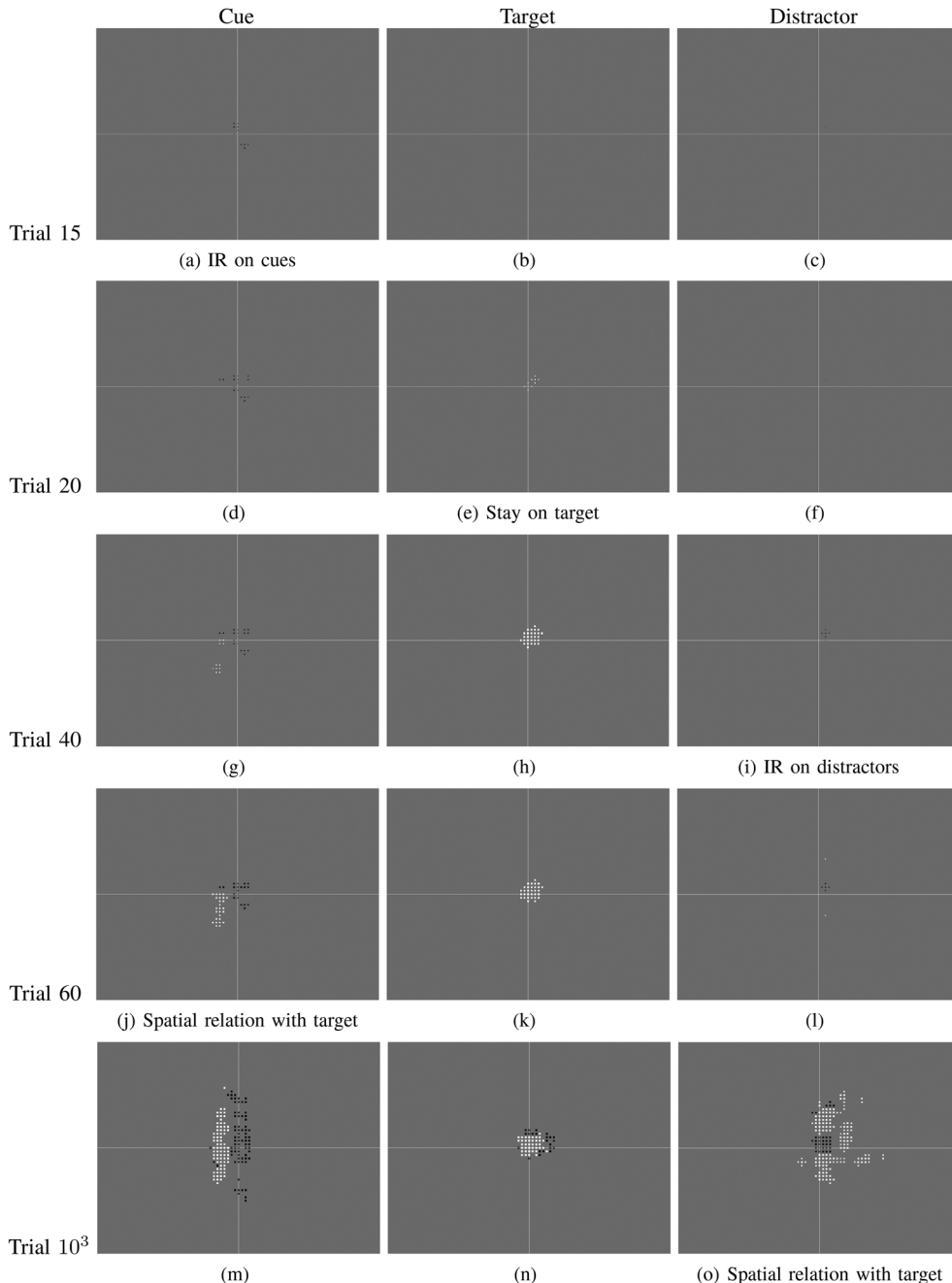
Fig. 8. Development of the vote map during 1000 training trials. The trial number reported on the left marks the critical developmental phases of the system indicated under the graphs (IR: inhibition of return). The three columns of graphs refer to the vote map activation when the system foveates a cue, the target, or a distractor. Within each graph, black and white dots indicate map units activated, respectively, below and above 0.5 (these values become negative/positive when sent to the PAM), while their size is proportional to the activation of the units. Graphs of cue and distractor: reprinted with permission from [80], Copyright IEEE (2010).

agrees with the Dynamic System Theory of Development for which it is not necessary nor useful that the phases of learning are strongly coded in the system as they naturally emerge from the interaction of the learning capacities of the system with the body and environment features [104], [105]. Notwithstanding the possible variability of child development, however, we think that some features of the developmental trajectory exhibited by BITPIC capture general processes that we expect to observe also in children. These features, which represent predictions of the model, are summarized in Section V.

D. The Potential Action Map Component

We have seen in Section I that the restricted scope of the fovea, although very beneficial to decrease the needed computations, gates away a lot of information and so augments the aliasing problem [113] affecting autonomous agents. The PAM allows BITPIC to face the aliasing problem on the basis of memory. An important feature of this memory, that distinguishes it from other memory-based solutions (e.g., [113]), is that it stores information on *past potential actions* rather

TABLE II
BASE, NEGPAM, AND PAM (WITH 0.1 AND 05 NOISE) SYSTEMS: FINAL
PERFORMANCE, AVERAGE NUMBER OF SACCADES ON TRIALS 450–500,
NUMBER OF TRIALS TO ACHIEVE A STEADY PERFORMANCE (I.E., THE
REACHING OF TARGET IN 7 SACCADES), AVERAGE NUMBER OF SACCADES ON
TRIALS 1–250, AND MAXIMUM NUMBER OF SACCADES PER TRIAL

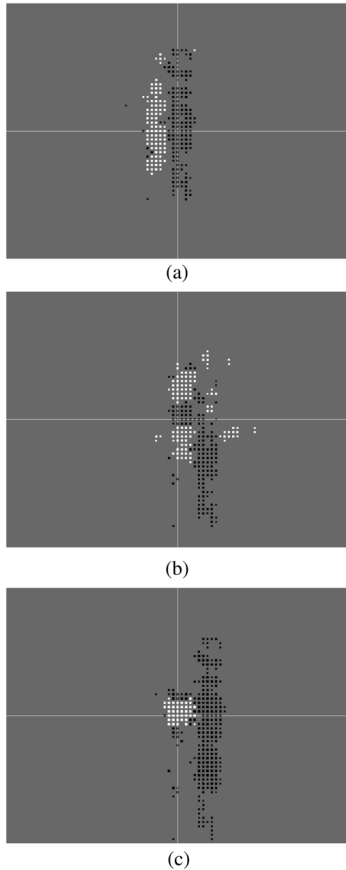|	Final perf.	N. sac. 450-500	Trials steady	N. sac. 1-250	Max sac. per trial
BASE	0.96	6.63	314	29.80	153
NEGPAM	0.97	6.42	88	17.55	263
PAM (0.5)	0.98	6.44	31	8.39	130
PAM (0.1)	0.98	6.75	34	10.25	75

Fig. 9. Activation of the PAM in a sequence of three saccades targeting (a) a cue, (b) a distractor, and (c) the target, in the scene of Fig. 1(a). Black and white dots indicate units with a negative and positive activation, respectively, while their size is proportional to the absolute activation of the units.

than on *past percepts*. In particular, for each foveated item the vote map estimates the best location in space to find relevant items, and this information is stored in the PAM and used for the current and future saccades. An action-based memory has several advantages based on the fact that saccade actions have a very low dimensionality (pan and tilt) that allow the system to conflate the information collected up to a certain moment into a 2D map. This also allows the PAM to be unaffected by the curse of dimensionality problem, contrary to perception-based memories that need to store different possible sequences of objects.

The PAM component ameliorates the aliasing problem under several aspects. First, it generates an inhibition of return by generating a repulsion to look again at specific objects. The reasons

is that staying with the eye on a particular object incurs in a small cost and is also not advantageous with respect to looking to targets or other relevant cues, so it gets penalized with learning (see also [76]). The resulting repulsion developed by the vote map is stored in the PAM and thus is available to the system in the following saccades even when it is looking at other objects. This prevents the system from moving back to previously foveated objects if this is not useful. Importantly, and differently from other system (e.g., [51] and [76]), this inhibition of return is emergent, i.e., learned, rather than hardwired. This allows the system to tune at best the repulsion (i.e., the negativity of the vote map) given the variable bottom–up component drive and the specific task needs.

The PAM also generates *positive biases* to look with higher chances to promising regions of space so overcoming various impairing aspects of perceptual aliasing. First, the PAM enhances exploration by allowing to explore several potential targets indicated by a cue without the need to look back to the latter, as a system without memory would do. Second, the PAM allows BITPIC to accumulate the bias to look at certain regions by looking at informative objects in sequence, so overcoming the impossibility of the fovea to integrate their biases by looking at them at the same time. Last, and most novel, we have shown that the PAM speeds up learning as it allows the transfer of the information on the spatial relations between relevant objects to other objects.

V. CONCLUSION AND FUTURE WORK

A key tenet of active vision is that the use of a movable fovea with a small size can drastically reduce the computational burden involved by visual sensors. However, this study highlighted that such solution generates four critical problems when applied to embodied and situated agents that interact with the environment in ecological conditions. We proposed a new system, called BITPIC, that *integrates* four bioinspired principles to solve the four problems: the autonomous acquisition of task-dependent visual skills with trial-and-error processes, a close coupling between attention and manipulation actions, a close interaction between top–down and bottom–up attention mechanisms, and finally an action-oriented memory. Based on this, we have also investigated the novel issue of how *active vision can enhance learning*.

The integration of the system principles has led to the emergence of notable properties of the system manifested within an interesting developmental trajectory. These properties represent predictions of the model that might be verified/falsified in future empirical experiments. We summarize the main ones as follows:

- The *adaptive processes of top–down attention*, integrated with the other ingredients of the system, lead to the *emergence of task-dependent, action-related, parsimonious visual routines* which support a *very efficient visual learning*.
- The *attention-manipulation coupling is essential for attention control learning* as it can support the *key transfer of value* from the realm of manipulation to task-relevant objects in the visual domain.
- *Bottom-up saliency greatly facilitates top–down attention learning by producing a structured visual exploration* of the environment (versus a random exploration typical of standard trial-and-error learning processes).
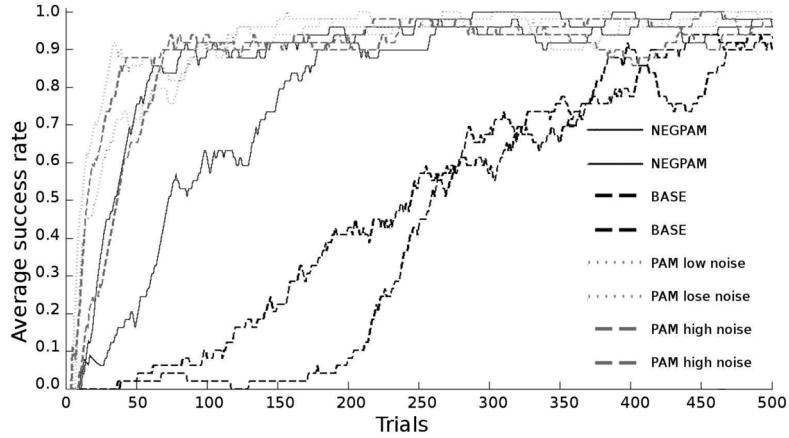
Fig. 10. Evolution of the success rate (proportion of trials ended with a successful reaching to the target, computed over a moving window of 20 trials) of the BASE, NEGPAM, and PAM systems during learning. The graph reports two replications of the simulation for each condition.
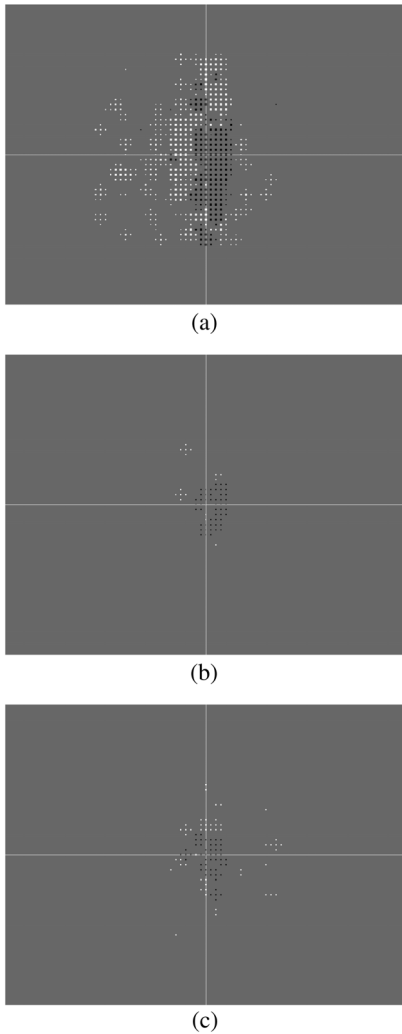

Fig. 11. Activation of the vote map of the BASE, PAM, and NEGPAM systems in the initial phases of learning, when they foveate the cue. Data plotted as in Fig. 8 (a) BASE (b) PAM (c) NEGPAM.

Fig. 12. Activation of the potential action memory map of the PAM and NEGPAM systems in the initial phases of learning when they foveate the target. Data plotted as in Fig. 9 (a) PAM. (b) NEGPAM.
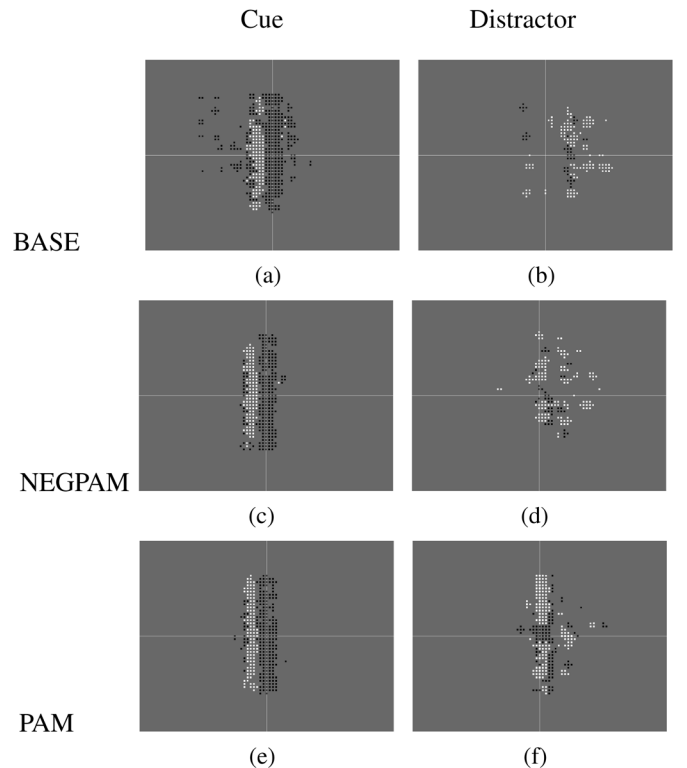
Fig. 13. Vote map of the BASE (a,b), NEGPAM (c,d), and PAM (e,f) systems after learning, when they foveate a cue (a,c,e) or a distractor (b,d,f). Data plotted as in Fig. 8.

- The *memory of potentially useful saccades ameliorates the aliasing problem* and, together with attention learning, leads to the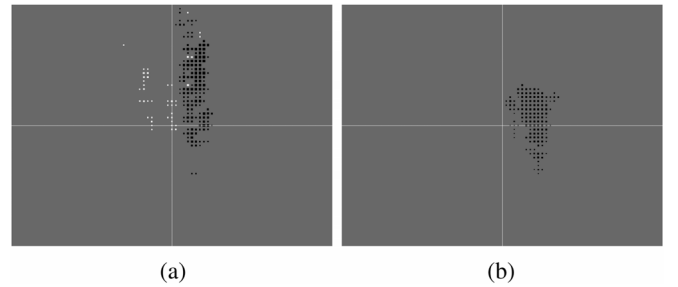 *emergence of a top–down flexible inhibition of return*; moreover, it *improves learning by transferring* spatial relation *knowledge* from object to object.

Although BITPIC was intendedly tested in a simple scenario to obtain a clear interpretation of the results, we expect that the main principles it is founded on are general and would extend to more complex scenarios although this would require strengthening low-level aspects of its components. We thus close the paper illustrating the possible challenges that we expect to encounter in scaling up the system to more realistic scenarios involving more complex objects, visual images, and noise, and we discuss possible actions to strengthen the model components to overcome such challenges.

a) Bottom-Up Attention Component–Possible Enhancements: Our experiences with both simulated and real cameras, with simple or realistic images, indicate that the bottom–up components based on filters as those proposed in [51], [52] are very robust to scene complexity and noise and allow systems to capture different saliency aspects of the scene elements without major difficulties. Some of these bottom–up filters can hence be added to the system as needed, e.g., to capture borders, corners, luminance intensity, or movement.

b) Top-Down Attention Component–Object Recognition: Objects need to be recognized in order to learn their spatial relations. Here we solved this problem by using color blobs, but object segmentation and recognition would be an important challenge using more complex objects. In this condition, a fovea larger than the one used here would be needed to support effective object recognition processes, e.g., based on SIFTs [62], sparse coding [83], or independent component analysis [18], [97]. Although current object recognition systems are still limited, they would nevertheless allow the system to achieve a satisfying level of performance with relatively complex scenes containing few objects.

c) Top-Down Attention Component–Relative/Absolute Reference Frames: The current top–down learning component uses a relative reference frame centred on the currently foveated object to decide potential saccade targets. This allows the system to capture relative spatial relations between objects but not absolute ones (e.g.: "whenever and wherever you see a toy, look at the face of your mother set at your right") [106]. Extensions of the model under test indicate that it is easy to use this alternative reference frame, e.g., by adding a second actor alongside the relative-reference actor used here, but how to best integrate the output of the two is an open issue.

d) Potential Action Map–a Robust Component: Experiments showed that the potential action map component is quite robust as it encodes information in a 2D space (the pan-tilt space) that does not depend on the complexity of the processed images [76]. Thus, the PAM could get input on where to look from relative and/or absolute top–down components, and from a bottom–up component, and encode information readily usable to control saccades in the "absolute" reference frame of the visuo-motor system, without being affected by the complexity of the input images.

e) Integration of Bottom-Up and Top-Down Attention Information–Superior Colliculus Functionalities and Brain Constraints: In the current system bottom–up and top–down information is integrated on the basis of a simple summation and the decision where to look is based on a winner-take-all process or a simple neural dynamic competition. In the brain, such integration and decision rely heavily on an important component

of the oculomotor control system, the superior colliculus. Future work will aim to implement more realistic versions of the superior colliculus functionalities and its interplay with basal ganglia and cortex, in particular the sophisticated architecture with which this system integrates different sources of information and the dynamic competition it employs to decide where to look [46], [98]. This would give enhanced properties to the current system, for example a staged and tunable integration of different information sources (here rigid). Moreover, it would allow a higher robustness to real-image noise as decisions would be taken by integrating information in time on the basis of a dynamic competition with parameters tunable on the fly as it happens in basal ganglia, rather than fixed as here [40]. In this respect, the amount of competition between competing bottom–up information sources, possibly based on parietal cortex (which represents an important source of input to superior colliculus and the basal ganglia areas controlling eye movements, [46]) has been suggested to play an important role in the development of attention [92]. In this respect, additional efforts will be spent to constrain the model architecture and functioning with general neuroscietific knowledge [28], [94] and neuroscientific knowledge related to development [55].

f) Motor System–Future Developments and New Challenges for Attention: The possibilities of enhancing the motor components of the system are several as motor control involves a whole set of challenges on its own, for example to implement reaching, grasping, obstacle avoidance, sophisticated movement trajectories, cyclic movements, multiple movements to solve different tasks [8]. We might for example expand the system manipulation capabilities on the basis of other models that we proposed within a developmental framework to capture learning of discrete or rhythmic movements [20], [24], [45], [82]. Note that a limitation of the current model is that the simplicity of manipulation behavior, consisting in simple reaching actions, reduces the relevance of the motor system which is simply triggered when the attention system fixates a point for few cycles. Notwithstanding this limitation, the presence of the motor system is important for a number of reasons. First, it supports the "attention for action" perspective in concrete modelling terms rather than in an abstract way. Second, the presence of the arm poses interesting challenges for the attentional system when the model is tested with the real robot as the arm becomes part of the visual image [76]. Last, more sophisticated manipulation actions as those mentioned above will open up a number of interesting new challenges for attention control as this would be called to guide manipulation in more sophisticated ways. For example, experiments show how while reaching a target by moving the hand around an obstacle human participants consistently foveate the obstacle before the target [44] so that the eye can inform the hand on the position of the obstacle to avoid.

g) Attention-Manipulation Coupling–Emergent Rather than Hardwired: A limitation of the current system involves the attention-manipulation coupling. Here we showed that, when present, this coupling has important *effects* on the system learning and behavior. However, the system does not show the *origin* of such coupling as this was hardwired. The interesting developmental problem of the coupling formation could be studied by having a manipulation system that learns

TABLE III
PARAMETER VALUES OF THE MODEL, WITH THE INDICATION OF THE FORMULA WHERE THEY ARE EMPLOYED AND THEIR FUNCTION

Formula	Parameter	Value	Function of parameter
Formula 3	r_s	$+1.0$	Reward for successful reaching
Formula 3	r_f	-0.1	Reward for failing reaching
Formula 3	r_e	-0.005	Reward for a saccade
Formula 3	γ	0.9	Discount factor
Formula 4	η_c	10^{-7}	Critic learning rate
Formula 5	η_a	10^{-5}	Actor learning rate
Formula 6	δ_{pam}	0.7	PAM decay
Formula 7	δ_{sm}	0.9	Saliency map decay
Formula 7	th_{sm}	0.7	Saliency map triggering threshold
Formula 8	δ_{apm}	0.9	Arm posture map decay
Formula 8	th_{apm}	0.3	Arm movement triggering threshold

concurrently with the attention system, rather than before it as done here. This modification might lead to the emergence of very interesting developmental trajectories given the reciprocal dependency of the visual and manipulation components, with the former furnishing information to the latter, and the latter procuring the learning feedback for the former.

h) Reward Learning Signals–Extrinsic and Intrinsic Motivations: A last aspect of the model that might be enhanced involves the generation of the positive reward signals, here delivered after a successful touch of a target object. Motivations are paramount for development [111]. In particular, motivations play a key role in guiding trial-and-error learning on the basis of the generation of *extrinsic rewards*, supporting the accomplishment of biologically relevant resources and a successful homeostatic regulations, and *intrinsic rewards*, related to the acquisition of new knowledge and skills [6]. The enhancement of the system with reward generators based on intrinsic motivations would be important to support a truly autonomous open-ended development of the system [9], [14], [15], [84]: this might lead to discover further general principles of the development of the visual system [91] and the coupled attentional-manipulation system [7], [101].

i) Validation of the Model Against Behavioral Data: Once the model has been enhanced in some of these ways, it should also be tested against specific behavioral data, for example to verify the type of predictions illustrated at the beginning of this section. As an example, consider [68] where a simplified version of BITPIC was successfully used to reproduce and explain the anticipatory looking behavior of infants engaged in exploring a computer image that could change depending on where the infants looked.

Overall, these considerations suggest that the architectural and functioning principles of BITPIC are very flexible and can be extended in multiple ways to address multiple interesting empirical and theoretical issues.

APPENDIX
COMPUTATIONAL DETAILS ON THE SYSTEM

A. Attention Control Components

In the following, bold symbols represent column vectors, while bold capital letters represent matrices. The parameters of the model are summarized in Table III (see also Table I).

Input Images: The input images contain colored rectangles (red, green, or blue) in a black background. The images appear in a CRT (Cathode Ray Tube) monitor whose display is used

as the working plane of the arm (Fig. 2(a)). The camera of the system looks down to the working plane and the image it captures overlaps with the monitor image. The 240×320 pixel RGB image captured by the camera is overlapped onto a black background to obtain a larger "scene-image" explored by the "virtual" moving camera of the system. The image of the virtual camera is obtained by cutting a 480×640 image from the scene-image and is used as input sent to the attention controller (Fig. 2(b)). In this way, all objects in the whole monitor image are always seen by the system, but their position within the input image depends on the system gaze direction. This gaze direction can fall in any point of the monitor image (and never outside it), so the system can foveate any object in the monitor. In the simulated tests presented here the image sent to the monitor in the real setup is directly used to form the scene-image.

Periphery Map (Bottom-Up Saliency Map): This component forms a very simple bottom–up saliency map based on the presence of any type of color blob in the black background of the image. The component is formed by a map of 60×80 units, encoded in the vector **pm**, activated on the basis of 60×80 blocks of 8×8 pixels each: each unit is activated by averaging the RGB color values of the pixels forming the corresponding block to obtain one "grey value."

Actor-Critic Component (Top-Down Attention): The basic system has a very simple fovea formed by 12 units, encoded in the vector **f**, activated on the basis of 2×2 RGB pixels taken from the foveated point of the input image. The fovea image is fed into two feedforward neural networks forming a reinforcement-learning actor-critic architecture [99]. The *actor* is a neural network whose output layer is formed by a *vote map* of 60×80 sigmoid neurons encoded in the 4800-element vector **vm**. The weights of the actor form a matrix, encoded with $\mathbf{W}_a$, having 12×4800 elements. The fovea, **f**, activates the vote map, **vm**, as follows:

$$\mathbf{vm} = g\left[\mathbf{W}_a \mathbf{f}\right]$$
$$g[x] = 1/(1 + e^{-x}) \tag{1}$$

where $g[x]$ is a Sigmoid function.

The *critic* is a neural network with connection weights encoded in the 12-element vector $\mathbf{w}_c$ and a linear output unit v that activates as follows on the basis of the current fovea image:

$$v = \mathbf{w}_c^T \mathbf{f} \tag{2}$$

where T is the transpose operator.

The critic learns to evaluate the current fovea image on the basis of the expected future discounted rewards. The system gets a reward $r_s = 1$ after the execution of a successful reaching action that touches the red target object with the "hand" (i.e., the arm tip gets on it), and a reward $r_f = -0.1$ if it fails (representing an energy cost): these rewards are received only when the system accomplishes a reaching action. An additional reward of $r_e = -0.005$ is delivered at each step (representing the energy cost of the saccade). The reward, together with v, is used to compute the *TD-error* δ_t at time t [99]

$$\delta_t = ((r_{st} + r_{ft} + r_{et}) + \gamma v_t) - v_{t-1} \qquad (3)$$

where γ ($\gamma = 0.9$) is a discount factor decreasing the importance of rewards far in the future.

The TD-error is used to update the connection weights of the critic [99]

$$\mathbf{w}_{ct} = \mathbf{w}_{ct-1} + \eta_c \delta_t \mathbf{f}_{t-1} \qquad (4)$$

were η_c is a learning rate ($\eta_c = 10^{-7}$). This learning rule is applied after each saccade at each cycle of the simulation.

The actor weights $\mathbf{W}_a$ are trained on the basis of a Hebbian learning rule involving the weights linking the units of the vote map ($\mathbf{vm}_{t-1}$) corresponding to the active units of the saliency map $\mathbf{sm}_{t-1}$ (encoding the last saccade location, see below), and the active units of the last fovea image $\mathbf{f}_{t-1}$, modulated by the TD-error δ_t [76]

$$\mathbf{W}_{at} = \mathbf{W}_{at-1} + \eta_a \delta_t (\mathbf{sm}_{t-1} \bullet \mathbf{d}) \mathbf{f}_{t-1}^T$$
$$\mathbf{d} = (\mathbf{vm}_{t-1} \bullet (1 - \mathbf{vm}_{t-1})) \qquad (5)$$

where $\bullet$ is the elementwise product operator, $(\mathbf{vm}_{t-1} \bullet (1 - \mathbf{vm}_{t-1}))$ is the derivative of the Sigmoid function, and η_a is a learning rate ($\eta_a = 10^{-5}$). The learning rule is applied after each saccade at each cycle of the simulation. This formula implies that the connection weights between the active units of the fovea ($\mathbf{f}_t$) and the units of the vote map ($\mathbf{vm}_{t-1}$) corresponding to the active units of the saliency map ($\mathbf{sm}_{t-1}$, encoding the performed saccade) are increased if $\delta_t > 0$ and decreased if $\delta_t < 0$. The rule allows the system to learn the spatial relations between objects, for example to look at the left of a cue if there it can obtain a positive δ_t (e.g., because it finds the rewarding target).

Potential Action Map (Top-Down Attention Memory): The PAM, encoded in the 4800-element vector $\mathbf{pam}$, is a neural map formed by 60×80 leaky neurons that accumulate the information received from the vote map $\mathbf{vm}$ via one-to-one connections. Importantly, the signals sent by the vote map to the PAM, having initial values of 0.5 and ranging in [0, 1] due to the Sigmoid transfer function of the actor units, are scaled to [-0.5, +0.5] so that they can either activate or inhibit the PAM units. This allows the PAM to have negative activations and so, when necessary, to inhibit useless tendencies to look at highly salient objects signalled by the bottom–up saliency map to the overall saliency map selecting the saccade targets. Similarly important, during each saccade the PAM activation is shifted in the direction opposite to the eye motion so as to maintain coherent

eye-centred representations and to allow its content to be suitably integrated with the next activation of the vote map. Formally, the PAM is activated as follows:

$$\mathbf{pam}_t = f[\delta_{pam} \cdot \mathbf{pam}_s + (\mathbf{vm}_t - 0.5)]$$
$$\mathbf{pam}_s = pam[\mathbf{epm}_t, \mathbf{epm}_{t-1}, \mathbf{pam}_{t-1}] \qquad (6)$$

where δ_{pam} ($\delta_{pam} = 0.7$) is a decay factor, $pam[\mathbf{epm}_t, \mathbf{epm}_{t-1}, \mathbf{pam}_{t-1}]$ is a function returning the activity $\mathbf{pam}_s$ of the PAM shifted on the basis of the current and past eye postures, respectively, $\mathbf{epm}_t$ and $\mathbf{epm}_{t-1}$ (see below), and $f[x]$ is a Sigmoid function ranging in [-1, +1]. The PAM is reset at the beginning of each trial when the scene changes.

Saliency Map: This component, not to be confounded with the bottom–up saliency map, is formed by 60×80 units encoded in the 4800-element vector $\mathbf{sm}$. The map integrates the bottom–up information received via one-to-one connections from $\mathbf{pm}$ and the top–down information received via one-to-one connections from $\mathbf{pam}$ to select the next saccade target. Each unit of the map has a preferred saccade target established by overlapping the whole map with the 2D space formed by the possible gaze targets (recall that this space corresponds to the monitor working space). The map integrates information and selects a saccade target through a dynamic neural-field competition

$$\mathbf{sm}_t = \max[\delta_{sm} \cdot \mathbf{sm}_{t-1} + \mathbf{W}_{sm}\mathbf{sm}_{t-1}$$
$$+ \mathbf{pm}_t + \mathbf{pam}_t + \mathbf{n}_t, \; 0] \qquad (7)$$

where δ_{sm} ($\delta_{sm} = 0.9$) is a decay factor, $\mathbf{W}_{sm}$ is a 4800×4800 matrix encoding prewired lateral connection weights (these are close-excitatory and far-inhibitory connections depending on the distance between neurons as in dynamic field networks, [36]), $\mathbf{n}_t$ is a noise added to the map ($\mathbf{n}_t$ elements are randomly drawn from $[-0.1, +0.1]$ and have a different value for each unit at each time step), and $\max[.,.]$ is a function returning the largest value of its two arguments. The saccade is executed when a unit activation achieves a given threshold th_{sm} ($th_{sm} = 0.7$). When this happens, the eye pan-tilt displacement $(\Delta x, \Delta y)$ is computed on the basis of the desired saccade position determined by the weighted average of the preferred position of the units of $\mathbf{sm}$, with weights equal to their activation, and of the current eye posture stored in the eye posture map $\mathbf{epm}$ (see below).

In the experiments reported in this work, the $\mathbf{sm}$ dynamic neural competition was actually approximated by selecting the unit with the maximum activation within the map ("winning unit"; to this purpose, the units of the map were activated on the basis of Formula 7 by setting δ_{sm} to zero and $\mathbf{W}_{sm}$ to a matrix of zeros). After the winning unit was selected, the units of the map were activated on the basis of a Gaussian function centred on it to mimic the activation of the map at the end of the dynamic competition (the activation of the map was also used to train the actor as illustrated above). This approximation did not cause relevant behavioral differences with respect to actual dynamic competition and at the same time substantially accelerated the simulations.

B. Arm Control Components

The arm controller decides when to trigger a reaching action and executes it by moving the arm "hand" to the current gaze fixation point on the working plane. To this purpose, the controller performs an inverse kinematic mapping from the current eye posture (gaze fixation point) to the corresponding arm angles that allow the arm hand to be on the target, and then sends this desired posture to the servos of the robot. The functioning of the arm controller components and then their training are now explained more in detail.

Eye Posture Map: This component is a 30×40 neural map encoded in the 1200-element vector **epm**. The map encodes the current eye posture, i.e., the current pan and tilt angles of the moving camera, with a Gaussian population code centred on the posture.

Arm Posture Map: This 40×40 map, denoted with the vector **apm**, is the output layer of a neural network pretrained with a Kohonen algorithm during a phase of "motor babbling" illustrated below. During this training the map learns to encode in 2D the arm postures corresponding to the four angles of the arm joints (see below and [82]). During functioning, the map implements a neural competition to select a desired arm posture for the reaching action

$$\mathbf{apm}_t = \max \left[\delta_{apm} \cdot \mathbf{apm}_{t-1} + \mathbf{W}_{apm}\mathbf{apm}_{t-1} \right. \\ \left. + \mathbf{W}_{epm-apm}\mathbf{epm}_t, 0 \right] \tag{8}$$

where δ_{apm} ($\delta_{apm} = 0.9$) is a decay factor, $\mathbf{W}_{apm}$ is a 1600×1600 matrix encoding the weights of lateral close-excitatory far-inhibitory connections (similarly to **sm**), and $\mathbf{W}_{epm-apm}$ is a 1600×1200 matrix encoding the weights of the connections received from the eye posture map. The competition terminates and the action is executed when any unit of the map reaches a given threshold th_{apm} ($th = 0.3$). The reading out of the map, encoding the selected desired hand position, is performed by weight-averaging the preferred hand positions of the map units with weights corresponding to their activation. The arm posture map is reset at the beginning of each trial as the scene changes.

Arm Posture Readout Layer: This is a vector, denoted with **aprl**, of four sigmoid units encoding the desired arm joint angles corresponding to the desired hand position selected by **apm**. The layer is activated by the arm posture map through connections whose weights are encoded in a 4×1600 matrix $\mathbf{W}_{apm-aprl}$. The desired arm joint angles are sent to the four simulated servos of the arm and these drive the arm to the desired angles on the basis of Proportional Derivative (PD) controllers [93].

Training: The weights $\mathbf{W}_{epm-apm}$ and $\mathbf{W}_{apm-aprl}$ are trained on the basis of a *motor babbling* process [20] during which the arm performs random movements while the eye gaze direction is kept on the arm "hand." The random arm movements are performed so that the last of the three arm segments is kept parallel to the working plane and at a fixed distance from it (see Fig. 2(a)). This avoids problems related to redundant degrees of freedom (see [45] for a version of the motor component addressing redundancy issues). Training is composed of three learning phases executed in sequence (see [82] for further details): (a) the arm posture map (**apm**) learns to encode in 2D the arm postures (four angles) using a

Kohonen algorithm [59]: this process leads the map to extract the two main components of the posture; (b) the system learns the inverse kinematic mapping ($\mathbf{W}_{epm-apm}$) associating the eye gaze directions corresponding to the hand, encoded in the eye posture map (**epm**), with the corresponding arm postures, encoded in the arm posture map (**apm**), using a supervised learning algorithm (delta rule; [114]); (c) the system learns the mapping ($\mathbf{W}_{apm-aprl}$) from the hand position, encoded in the arm posture map (**apm**), to the desired arm angles, encoded in the arm posture readout layer (**aprl**), again on the basis of a supervised learning algorithm (delta rule).

After the three training phases have been accomplished, each gaze position fuels the accumulation of activation within the arm posture map. If the system foveates the same position for a long time, the arm posture map accumulates an activity that overcomes the threshold, so triggering the reaching action and at the same time defining its target. This leads the arm posture readout layer to issue desired arm angles to the arm servos that implement the actual movement.

REFERENCES

[1] A. Allport, "Selection for action: Some behavioral and neurophysiological considerations of attention and action," *Perspect. Percep. Action*, vol. 15, pp. 395–419, 1987.

[2] A. Allport, "Visual attention," in *Foundations of Cognitive Science*, M. Posner, Ed. Cambridge, MA: MIT Press, 1990, pp. 631–682.

[3] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *Int. J. Comput. Vision*, vol. 1, no. 4, pp. 333–356, Jan. 1988.

[4] A. Andreopoulos and J. K. Tsotsos, "A computational learning theory of active object recognition under uncertainty," *Int. J. Comput. Vision*, vol. 101, no. 1, pp. 95–142, 2013.

[5] R. Bajcsy, "Active perception," in *Proc. IEEE*, 1988, vol. 76, no. 8, pp. 966–1005.

[6] G. Baldassarre, "What are intrinsic motivations? A biological perspective," A. Cangelosi, J. Triesch, I. Fasel, K. Rohlfing, F. Nori, P. Y. Oudeyer, M. Schlesinger, and Y. Nagai, Eds., Proc. Int. Conf. Develop. Learn. Epigenetic Robot. (ICDL-EpiRob-2011) pp. E1–E8, IEEE New York, 2011.

[7] G. Baldassarre, 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 Netw.*, vol. 41, pp. 168–187, 2013.

[8] *Computational and Robotic Models of the Hierarchical Organisation of Behaviour*, G. Baldassarre and M. Mirolli, Eds. Berlin: Springer-Verlag, 2013.

[9] *Intrinsically Motivated Learning in Natural and Artificial Systems*, G. Baldassarre and M. Mirolli, Eds. Berlin: Springer, 2013.

[10] C. Balkenius, "Attention, habituation and conditioning: Toward a computational model," *Cogn. Sci. Quart.*, vol. 1, no. 2, pp. 171–204, 2000.

[11] C. Balkenius and B. Johansson, "Anticipatory models in gaze control: A developmental model," *Cogn. Process.*, Apr. 2007.

[12] D. H. Ballard, M. M. Hayhoe, P. K. Pook, and R. P. Rao, "Deictic codes for the embodiment of cognition," *Behav. Brain Sci.*, vol. 20, no. 4, pp. 723–42, 1997.

[13] D. H. Ballard, "Animate vision," *AI*, vol. 48, pp. 57–86, 1991.

[14] A. Barto, M. Mirolli, and G. Baldassarre, "Novelty or surprise?," *Frontiers in Psychol. – Cogn. Sci.*, vol. 4, no. 907, pp. 1–15, 2013.

[15] A. G. Barto, S. Singh, and N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills," J. Triesch and T. Jebara, Eds., Int. Conf. Develop. Learn. (ICDL2004) pp. 112–119, IEEE, 2004, UCSD Inst. Neural Comput. LaJolla, CA.

[16] J. Baxter and P. L. Bartlett, "Reinforcement learning in pomdp's via direct gradient ascent," in *ICML*, 2000, pp. 41–48.

[17] L. Berthouze and Y. Kuniyoshi, "Emergence and categorization of coordinated visual behavior through embodied interaction," *Auton. Robots*, vol. 5, no. 3-4, pp. 369–379, 1998.

[18] N. Bruce and J. Tsotsos, "Saliency based on information maximization," *Adv. Neural Inf. Process. Syst.*, vol. 18, pp. 155–162, 2006.

[19] N. J. Butko and J. R. Movellan, "Infomax control of eye movements," *IEEE Trans. Autonom. Mental Develop.*, vol. 2, no. 2, pp. 91–107, 2010.

[20] D. Caligiore, T. Ferrauto, D. Parisi, N. Accornero, M. Capozza, and G. Baldassarre, "Using motor babbling and hebb rules for modeling the development of reaching with obstacles and grasping," R. Dillmann, C. Maloney, G. Sandini, T. Asfour, G. Cheng, G. Metta, and A. Ude, Eds., International Conf. Cognitive Systems (CogSys2008) pp. E1–E8, Univ. Karlsruhe   Karlsruhe, Germany, Apr. 2–4, 2008, 2008.

[21] R. Caminiti, S. Ferraina, and A. B. Mayer, "Visuomotor transformations: Early cortical mechanisms of reaching," *Curr. Opin. Neurobiol.*, vol. 8, no. 6, pp. 753–761, 1998.

[22] S. Chikkerur, T. Serre, C. Tan, and T. Poggio, "What and where: A Bayesian inference theory of attention," *Vision Res.*, vol. 50, no. 22, pp. 2233–2247, 2010.

[23] P. S. Churchland, V. S. Ramachandran, and T. J. Sejnowski, "A critique of pure vision," in *Large-Scale Neuronal Theories of the Brain*, C. Koch and J. L. Davis, Eds.   Cambridge, MA, USA: MIT Press, 1994, pp. 23–60.

[24] A. L. Ciancio, L. Zollo, E. Guglielmelli, D. Caligiore, and G. Baldassarre, "Hierarchical reinforcement learning and central pattern generators for modeling the development of rhythmic manipulation skills," in A. Cangelosi, J. Triesch, I. Fasel, K. Rohlfing, F. Nori, P. Y. Oudeyer, M. Schlesinger, and Y. Nagai, Eds., Proc. 1st Joint IEEE Int. Conf. Develop. Learn. Epigenetic Robot. (ICDL-EPIROB 2011) pp. E1–E8, IEEE   New York, Aug. 24-27, 2011.

[25] P. Cisek, "Integrated neural processes for defining potential actions and deciding between them: A computational model," *J. Neurosci.*, vol. 26, no. 38, pp. 9761–9770, 2006.

[26] P. Cisek and J. F. Kalaska, "Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action?," *Neuron*, vol. 45, no. 5, pp. 801–814, 2005.

[27] D. Cliff and J. Noble, "Knowledge-based vision and simple visual machines," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 352, no. 1358, pp. 1165–1175, 1997.

[28] A. Cope and K. Gurney, "A biologically based model of active vision," in *Proc. Soc. Study of Artif. Intell. Simul. Behav. (AISB2012)*, 2012, The Soc. Study of Artif. Intell. Simul. Behav..

[29] D. Corbetta, Y. Guan, and J. L. Williams, "Visual attention and action selection in infant reaching," *Infancy*, vol. 7, no. 1, pp. 102–125, 2012.

[30] D. Corbetta, H. Dobbs, C. Ford, and Y. Guan, "The formation of perceptual-motor matching in infant reaching: A preliminary longitudinal study," *J. Sport Exercise Psychol.*, vol. 34, p. S158, 2012.

[31] C. E. Curtis, "Prefrontal and parietal contributions to spatial working memory," *Neurosci.*, vol. 139, no. 1, pp. 173–180, Apr. 2006.

[32] G. de Croon, E. O. Postma, and H. J. van den Herik, "A situated model for sensory-motor coordination in gaze control," *Pattern Recogn. Lett.*, vol. 27, no. 11, pp. 1181–1190, 2006.

[33] G. de Croon and E. O. Postma, "Sensory-motor coordination in object detection," in *Proc. IEEE Symp. Artif. Life (Ci- ALife 2007)*, 2007, pp. 147–154.

[34] Guido de Croon, Adaptive Active Vision Ph.D. thesis, Univ. Maastricht, 2008.

[35] C. W. Eriksen and J. D. St James, "Visual attention within and around the field of focal attention: A zoom lens model," *Percept. Psychophys.*, vol. 40, no. 4, pp. 225–240, 1986.

[36] W. Erlhagen and G. Schöner, "Dynamic field theory of movement preparation," *Psychol. Rev.*, vol. 109, no. 3, pp. 545–572, 2002.

[37] A. G. Feldman, "Once more on the equilibrium-point hypothesis (lambda model) for motor control," *J. Motor Behav.*, vol. 18, pp. 17–54, 1986.

[38] J. M. Findlay and I. D. Gilchrist, *Active vision: The psychology of looking and seeing*.   Oxford, U.K.: Oxford Univ. Press, 2003.

[39] J. M. Findlay and R. Walker, "How are saccades generated?," *Behav. Brain Sci.*, vol. 22, no. 04, pp. 706–713, 1999.

[40] V. G. Fiore, V. Sperati, F. Mannella, M. Mirolli, K. Gurney, K. Firston, R. J. Dolan, and G. Baldassarre, "Keep focussing: Striatal dopamine multiple functions resolved in a single mechanism tested in a simulated humanoid robot," *Front. Psychol.–Cogn. Sci.*, vol. 5, no. 124, 2014.

[41] K. Friston, R. A. Adams, L. Perrinet, and M. Breakspear, "Perceptions as hypotheses: Saccades as experiments," *Front. Psychol.*, vol. 3, p. 151, 2012.

[42] M. A. Goodale and A. D. Milner, "Separate visual pathways for perception and action," *Trends Neurosci.*, vol. 15, no. 1, pp. 20–25, 1992.

[43] J. P. Gottlieb, M. Kusunoki, and M. E. Goldberg, "The representation of visual salience in monkey parietal cortex," *Nature*, vol. 391, no. 6666, pp. 481–484, 1998.

[44] M. Hayhoe and D. Ballard, "Eye movements in natural behavior," *Trends Cogn. Sci.*, vol. 9, no. 4, pp. 188–194, 2005.

[45] O. Herbort, D. Ognibene, M. V. Butz, and G. Baldassarre, "Learning to select targets within targets in reaching tasks," in *Proc. IEEE 6th Int. Conf. Develop. Learn. (ICDL 2007)*, Jul. 11-13, 2007, pp. 7–12.

[46] 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.

[47] J. C. Houk, J. L. Adams, and A. G. Barto, "A model of how the basal ganglia generate and use neural signals that predict reinforcement," in *Models of Inf. Process. Basal Ganglia*, J. C. Houk, J. L. Davis, and D. G. Beiser, Eds.   Cambridge, MA: MIT Press, 1995, pp. 249–270.

[48] D. H. Hubel, *Eye, Brain, and Vision*.   New York: Scientif. Amer. Books, 1988.

[49] M. Hulse, S. McBride, J. Law, and M. Lee, "Integration of active vision and reaching from a developmental robotics perspective," *IEEE Trans. Autonom. Mental Develop.*, vol. 2, no. 4, pp. 355–367, 2010.

[50] A. Ikkai and C. E. Curtis, "Common neural mechanisms supporting spatial working memory, attention and motor intention," *Neuropsychol.*, vol. 49, no. 6, pp. 1428–1434, 2011.

[51] C. Itti and L. Koch, "Feature combination strategies for saliency-based visual attention systems," *J. Electron. Imag.*, vol. 10, no. 1, pp. 161–169, 2001.

[52] L. Itti, N. Dhavale, and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," B. Bosacchi, D. B. Fogel, and J. C. Bezdek, Eds., Proc. SPIE 48th Ann. Int. Symp. Opt. Sci. Technol. pp. 64–78   Bellingham, WA, 2003, SPIE.

[53] L. Itti and C. Koch, "Computational modelling of visual attention," *Nat. Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, 2001.

[54] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[55] M. H. Johnson, *Developmental Cognitive Neuroscience*.   Malden, MA: Wiley, 2010.

[56] D. Kirsh and P. Maglio, "On distinguishing epistemic from pragmatic action," *Cogn. Sci.*, vol. 18, no. 4, pp. 513–549, 1994.

[57] R. M. Klein, "Inhibition of return," *Trends Cogn. Sci.*, vol. 4, no. 4, pp. 138–147, 2000.

[58] C. Koch and S. Ullman, Selecting one among the many: A simple network implementing shifts in selective visual attention Mass. Inst. Technol., Cambridge, MA, 1984, Technical report.

[59] T. Kohonen, *Self-organizing maps*.   Berlin: Springer-Verlag, 1997.

[60] S. Kollmorgen, N. Nortmann, S. Schrder, and P. Knig, "Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention," *PLoS Comput. Biol.*, vol. 6, no. 5, p. e1000791, 2010.

[61] R. J. Krauzlis, M. A. Basso, and R. H. Wurtz, "Shared motor error for multiple eye movements," *Science*, vol. 276, no. 5319, pp. 1693–1695, 1997.

[62] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vision*, Sep. 20-27, 1999, vol. 2, pp. 1150–1157, 1999.

[63] X. Lu, M. Matsuzawa, and O. Hikosaka, "A neural correlate of oculomotor sequences in supplementary eye field," *Neuron*, vol. 34, no. 2, pp. 317–325, 2002.

[64] M. Lungarella and O. Sporns, "Information self-structuring: Key principle for learning and development," in *Proc. 4th Int. Conf. Develop. Learn.*, Osaka, Japan, Jul. 19-21, 2005, pp. 25–30, IEEE.

[65] G. L. Malcolm and J. M. Henderson, "Combining top–down processes to guide eye movements during real-world scene search," *J. Vision*, vol. 10, pp. 1–11, 2010.

[66] F. Mannella and G. Baldassarre, "A neural-network reinforcement-learning model of domestic chicks that learn to localize the centre of closed arenas," *Philosoph. Trans. R. Soc. B Biolog. Sci.*, vol. 362, no. 1479, pp. 383–401, 2007.

[67] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*.   New York: Freeman, 1982.

[68] R. Marraffa, V. Sperati, D. Caligiore, J. Triesch, and G. Baldassarre, "A bio-inspired attention model of anticipation in gaze-contingency experiments with infants," J. Movellan and M. Schlesinger, Eds., IEEE Int. Conf. Develop.and Learning and Epigenetic Robotics (ICDL-EpiRob-2012) pp. e1–e6   New York, 2012, IEEE, Nov. 7–9, 2012, San Diego CA.

[69] S. Minut and S. Mahadevan, "A reinforcement learning model of selective visual attention," J. P. Müller, E. Andre, S. Sen, and C. Frasson, Eds., Proc. Fifth International Conference on Autonomous Agents (AGENTS "01) pp. 457–464   New York, 2001, ACM.

[70] M. Mirolli, T. Ferrauto, and S. Nolfi, "Categorisation through evidence accumulation in an active vision system," Connect. Sci., vol. 22, no. 4, pp. 331–354, 2010.

[71] M. Mishkin, L. G. Ungerleider, and K. A. Macko, "Object vision and spatial vision: Two cortical pathways," Trends in Neurosci., vol. 6, pp. 414–417, 1983.

[72] K. Nakamura and C. L. Colby, "Updating of the visual representation in monkey striate and extrastriate cortex during saccades," Proc. Natl. Acad. Sci. USA, vol. 99, no. 6, pp. 4026–4031, 2002.

[73] S. Nolfi, "Power and the limits of reactive agents," Neurocomputing, vol. 42, no. 1-4, pp. 119–145, 2002.

[74] S. Nolfi and D. Floreano, Evolutionary Robotics: The Biology, Intelligence, and Technology.   Cambridge, MA: MIT Press, 2000.

[75] D. Ognibene, "Ecological Adaptive Perception from a Neuro-Robotic perspective: theory, architecture and experiments," Ph.D. thesis, Univ. Genoa, Genoa, Italy, 2009.

[76] D. Ognibene, C. Balkenius, and G. Baldassarre, "Integrating epistemic action (active vision) and pragmatic action (reaching): A neural architecture for camera-arm robots," M. Asada, J. Tani, J. Hallam, and J.-A. Meyer, Eds., The Tenth International Conference on the Simulation of Adaptive Behaviour (SAB2008) Berlin, 2008, Springer   Osaka, Japan, Jul. 7-12, 2008.

[77] D. Ognibene, C. Balkenius, and G. Baldassarre, "A reinforcement-learning model of top–down attention based on a potential-action map," in The Challenge of Anticipation–A Unifying Framework for the Analysis and Design of Artificial Cognitive Systems, ser. Lecture Notes in Artif. Intell., G. Pezzulo, M. V. Butz, C. Castelfranchi, and R. Falcone, Eds.   Berlin: Springer, 2008, vol. 5225, pp. 161–184.

[78] D. Ognibene, N. Catenacci Volpi, G. Pezzulo, and G. Baldassarre, "Learning epistemic actions in model-free memory-free reinforcement learning: Experiments with a neuro-robotic model," in Living Machines "13–Proc. 2nd Int. Conf. Biomim. Biohybrid Syst., Berlin, Jul.-Aug. 29-2, 2013, pp. 191–203.

[79] D. Ognibene and Y. Demiris, "Towards active event recognition," in Proc. 23rd Int. Joint Conf. Artif. Intell., 2013, pp. 2495–2501, AAAI Press.

[80] D. Ognibene, G. Pezzulo, and G. Baldassarre, "How can bottom–up information shape learning of top–down attention-control skills?," B. Kuipers, T. Shultz, A. Stoytchev, and C. Yu, Eds., Proc. 9th IEEE Int. Conf. Develop. Learn. (ICDL2010) pp. 231–237   New York, MA, IEEE, Univ. Michigan   Ann Arbor, MA, Aug. 18-21, 2010.

[81] D. Ognibene, G. Pezzulo, and G. Baldassarre, "Learning to look in different environments: An active-vision model which learns and readapts visual routines," S. Doncieux, G. Benoit, A. Guillot, J. Hallam, J.-A. Meyer, and J.-B. Mouret, Eds., From Animals to Animats 11: Proc. Eleventh International Conference on the Simulation of Adaptive Behavior (SAB 2010) volume 6226 of Lectures Notes in Computer Science pp. 199–210   Berlin, 2010, Springer.

[82] D. Ognibene, A. Rega, and G. Baldassarre, "A model of reaching that integrates reinforcement learning and population encoding of postures," S. Nolfi, G. Baldassarre, R. Calabretta, J. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, and D. Parisi, Eds., From Animals to Animats 9: Proc. 9th Int. Conf. Simulation of Adapt. Behav. (SAB 2006) pp. 381–393   Berlin, Rome, Italy, Springer, Sep. 25-29, 2006.

[83] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information," J. Neurosci., vol. 13, no. 11, pp. 4700–4719, 1993.

[84] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic motivation systems for autonomous mental development," IEEE Trans. Evolution. Comput., vol. 11, pp. 265–286, 2007.

[85] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom–up gaze allocation in natural images," Vision Res., vol. 45, no. 18, pp. 2397–2416, 2005.

[86] A. Pouget, P. Dayan, and R. Zemel, "Information processing with population codes," Nat. Rev. Neurosci., vol. 1, no. 2, pp. 125–132, 2000.

[87] R. P. N. Rao and O. Fuentes, "Hierarchical learning of navigational behaviors in an autonomous robot using a predictive sparse distributed memory," Mach. Learn., vol. 31, no. 1, pp. 87–113, 1998.

[88] F. Rea, G. Metta, and C. Bartolozzi, "Event-driven visual attention for the humanoid robot icub," Front Neurosci., vol. 7, p. 234, 2013.

[89] D. L. Robinson and S. E. Petersen, "The pulvinar and visual salience," Trends Neurosci., vol. 15, no. 4, pp. 127–132, 1992.

[90] C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe, "Task and context determine where you look," J. Vision, vol. 7, no. 14, pp. 16.1–1620, 2007.

[91] M. Schlesinger, "Investigating the origins of intrinsic motivation in human infants," in Intrinsically Motivated Learning in Natural and Artificial Systems, G. Baldassarre and M. Mirolli, Eds.   Berlin: Springer-Verlag, 2013, pp. 367–392.

[92] M. Schlesinger, D. Amso, and S. P. Johnson, "The neural basis for visual selective attention in young infants: A computational account," Adapt. Behav., vol. 15, no. 2, pp. 135–148, 2007.

[93] Modelling and control of robot manipulators, L. Sciavicco and B. Siciliano, Eds.   Berlin, Germany: Springer-Verlag, 2000.

[94] V. Sperati and G. Baldassarre, "Learning where to look with movement-based intrinsic motivations: A bio-inspired model," in Proc. 4th Joint IEEE Int. Conf. Develop. Learn. Epigenetic Robotics (ICDL-EpiRob2014), Genoa, Italy, Oct. 13–16, , pp. E1–E8.

[95] N. Sprague and D. H. Ballard, "Eye movements for reward maximization," in Advances in Neural Information Processing Systems 16, S. Thrun, L. Saul, and B. Schölkopf, Eds.   Cambridge, MA: MIT Press, 2004.

[96] M. W. Spratling, "Predictive coding as a model of the v1 saliency map hypothesis," Neural Netw., vol. 26, pp. 7–28, 2012.

[97] J. V. Stone, Independent Component Analysis.   Oxford: Wiley Online Library, 2004.

[98] V. Stuphorn, E. Bauswein, and K. P. Hoffmann, "Neurons in the primate superior colliculus coding for arm movements in gaze-related coordinates," J. Neurophysiol., vol. 83, no. 3, pp. 1283–1299, 2000.

[99] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction.   Cambridge, MA: MIT Press, 1998.

[100] M. Suzuki and D. Floreano, "Enactive robot vision," Adapt. Behav., vol. 16, no. 2-3, pp. 122–128, 2008.

[101] 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," Experiment. Brain Res., Apr. 2014.

[102] B. W. Tatler, M. M. Hayhoe, M. F. Land, and D. H. Ballard, "Eye guidance in natural vision: Reinterpreting salience," J. Vision, vol. 11, no. 5, pp. 1–23, 2011.

[103] J. B. Tenenbaum, T. L. Griffiths, and C. Kemp, "Theory-based Bayesian models of inductive learning and reasoning," Trends Cogn. Sci., vol. 10, no. 7, pp. 309–318, 2006.

[104] E. Thelen, "Developmental origins of motor coordination: Leg movements in human infants," Develop. Psychobiol., vol. 18, no. 1, pp. 1–22, 1985.

[105] E. Thelen and L. B. Smith, A dynamic systems approach to the development of cognition and action.   Cambridge, MA, USA: MIT Press, 1994.

[106] S. P. Tipper, L. A. Howard, and G. Houghton, "Action-based mechanisms of attention," Philos. Trans. R. Soc. London B. Biol. Sci., vol. 353, no. 1373, pp. 1385–1393, 1998.

[107] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," Cogn. Psychol., vol. 12, no. 1, pp. 97–136, 1980.

[108] S. Ullman, "Visual routines," Cognition, vol. 18, no. 1-3, pp. 97–159, 1984.

[109] P. F. M. J. Verschure, "Distributed adaptive control: A theory of the mind, brain, body nexus," Biologically Inspired Cogn. Architectures (BICA2012), vol. 1, pp. 55–72, 2012.

[110] P. F. M. J. Verschure, T. Voegtlin, and R. J. Douglas, "Environmentally mediated synergy between perception and behaviour in mobile robots," Nature, vol. 425, no. 6958, pp. 620–624, 2003.

[111] C. von Hofsten, "An action perspective on motor development," Trends Cogn. Sci., vol. 8, no. 6, pp. 266–272, 2004.

[112] Q. Wang, J. Bolhuis, T. Rothkopf, C. A. 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, p. e30884, 2012.

[113] S. D. Whitehead and D. H. Ballard, "Learning to perceive and act by trial and error," *Mach. Learn.*, vol. 7, no. 1, pp. 45–83, 1991.

[114] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *Neurocomputing: foundations of research*, pp. 123–134, 1988.

[115] A. L. Yarbus, *Eye Movements and Vision*. New York, NY, USA: Plenum, 1967.

[116] S. Zeki, *A Vision of the Brain*. Oxford, U.K.: Oxford Univ. Press, 1993.

Dimitri Ognibene received the Ph.D. degree in robotics from the University of Genoa, Italy, in 2009.

He was a Postdoctoral Research Associate at the Institute of Cognitive Sciences and Technologies, Italian Research Council, Rome, and at the Personal Robotics Laboratory, Imperial College London. Currently, he is a Postdoctoral Research Associate at the Center for Robotics Research, Department of Informatics, Kings College London (CoRe-KCL), Strand, London. He was also a Visiting Researcher at the Computer Science Department, University of Massachusetts, Amherst, and at the University of Reykjavik, Iceland. His research is centred on methods and principles for adaptive and active perception and computation in biological and artificial systems.

Dr. Ognibene was invited to speak at the 2013 International Symposium for Attention in Cognitive Systems, as well as in various neuroscience, robotics, and machine-learning labs. He has presented his work in several international conferences on artificial intelligence, adaptation, and development, and has published on international peer-reviewed journals. sites.google.com/site/dimitriognibenehomepage/

Gianluca Baldassare received the B.A. and M.A. degrees in economics in 1998, and in 1999, received the M.Sc. degree in cognitive psychology and neural networks from the University of Rome "La Sapienza". In 2003, he received the Ph.D. degree in computer science at the University of Essex, U.K. (research: "Planning with Neural Networks").

He is a Researcher at the Institute of Cognitive Sciences and Technologies, Italian National Research Council (ISTC-CNR), Rome. He was a postdoctoral researcher at ISTC-CNR working on swarm robotics. Since 2006, he has coordinated the Research Group "LOCEN Laboratory of Computational Embodied Neuroscience." During 2006-2009, he was Principal Investigator for ISTC-CNR within the EU project "ICEA – Integrating Cognition Emotion and Autonomy"; in 2009-2013 he was the Coordinator of the European Integrated Project "IM-CLeVeR – Intrinsically-Motivated Cumulative-Learning Versatile Robots." His research interests are on cumulative learning of sensorimotor skills driven by extrinsic and intrinsic motivations in animals, humans, and robots. He studies these topics with two approaches: (a) computational system-level models constrained by detailed data on brain and behavior and, when possible, embodied; (b) developmental robotic models employing both simulated and real robots. He has more than 100 international peer-review publications in psychology, neuroscience, and robotics conferences and journals. http://www.istc.cnr.it/people/gianluca-baldassarre