

AFFORDANCES AND COMPATIBILITY EFFECTS: A NEURAL-NETWORK COMPUTATIONAL MODEL

D. CALIGIORE*^{§+}, A. M. BORGHI^{§*}, D. PARISI^{*§}, G. BALDASSARRE*

**Consiglio Nazionale delle Ricerche, Istituto di Scienze e Tecnologie della Cognizione,
Via San Martino della Battaglia 44, I-00185 Roma, Italy
{daniele.caligiore, domenico.parisi, gianluca.baldassarre}@istc.cnr.it*

*+Università Campus Bio-Medico,
Via Alvaro del Portillo 21, I-00128 Roma, Italy*

*§Università di Bologna, Dipartimento di Psicologia,
Viale Berti Pichat 5, I-40127 Bologna, Italy
annamaria.borghi@unibo.it*

Behavioural and brain imaging evidence has shown that seeing objects automatically evokes “affordances”, for example it tends to activate internal representations related to the execution of precision or power grips. In line with this evidence, Tucker and Ellis [1] found a compatibility effect between object size (small and large) and the kind of grip (precision and power) used to respond whether seen objects were artefacts or natural objects. This work presents a neural-network model that suggests an interpretation of these experiments in agreement with a recent theory on the general functions of prefrontal cortex. Prefrontal cortex is seen as a source of top-down bias in the competition for behavioural expression of multiple neural pathways carrying different information. The model successfully reproduces the experimental results on compatibility effects and shows how, although such a bias allows organisms to perform actions which differ from those suggested by objects’ affordances, these still exert their influence on behaviour as reflected by longer reaction times.

1. Introduction

According to the traditional view of cognition, perception precedes action and is not influenced by it. Sensory stimuli determine how the world is represented in organisms’ nervous systems whereas processes underlying actions play only a role on how they intervene on the environment to modify it. This passive view of knowledge is challenged by recent behavioural [2], physiological [3] and brain imaging [4] evidence showing that organisms’ internal representations of the world depend on the actions with which they respond to sensory stimuli.

In this perspective, the notion of affordance [5] has been given new relevance. An affordance is a quality of an object which is directly accessible to

an organism and suggests its possible interactions, uses and actions. Many works provide evidence in favour of an automatic activation of affordances during the observation of objects [6][7].

One way of studying how internal representations of objects and concepts rely upon motor information is to devise experimental tasks in which participants are shown objects and are asked to produce actions which are either in agreement (congruent trials) or in contrast (incongruent trials) with the actions typically associated with those objects (e.g., to grasp an object with the appropriate grip). As objects automatically elicit the activation of their related affordances, if participants find it more difficult (e.g. as revealed by longer reaction times) to act in incongruent trials than in congruent ones, one can infer that objects are at least in part represented in terms of potential actions.

Tucker and Ellis [1] performed an experiment with this compatibility paradigm. Participants were asked to classify large and small objects into artefacts or natural categories by mimicking either a precision or a power grip while acting on a customised joystick. Importantly, object size was not relevant to the categorisation task. The authors found a compatibility effect between object size (large and small) and motor response (power and precision grip), namely shorter reaction times (RTs) in congruent trials than in incongruent ones. These results show that object knowledge relies not only on objects perceptual features but also on the actions that can be performed on them.

This work presents a bio-mimetic neural-network model which allows interpreting the results of the aforementioned experiments on the basis of the integration of three general principles of brain functioning. The first regards the broad organization of brain cortex underlying visual processing into the “dorsal and ventral streams” [25]. The ventral stream is a neural pathway which carries information, among other things, about the identity of objects (“what”). The dorsal stream is a neural pathway which carries spatial information, for example about the shape and location of objects (“where/how”). This pathway implements the “affordances” of objects which can be learned during the first months of life, but also in the rest of life, on the basis of spontaneous environment explorations. The second principle concerns the general theory on the functions of prefrontal cortex (PFC) recently proposed by Miller and Cohen [8]. This theory views PFC as an important source of top-down biasing where different neural pathways carrying different types of information compete for expression in behaviour [8][24]. Finally, the third principle is about the use of neural networks based on the dynamic field approach [21] which allow accounting for reaction times on the basis of biologically plausible mechanisms.

In agreement with the computational neuroscience approach [9], the model is not only requested to reproduce behaviours observed in experiments but it is also constrained, at the level of the model overall architecture and functioning, by the known anatomy and physiology of brain structures underlying the behaviours investigated [10].

2. The model

2.1. Simulated robotic set-up

The model controls a simulated 3D artificial organism endowed with a visual system, a human-like 3-Segments/4-DOF arm, and a 21-Segments/19-DOF hand (Fig. 1a). The visual system is formed by a simulated “eye” (a 630×630 pixel RGB camera with a 120° pan angle and a 120° tilt angle) mounted 25cm above the arm’s “shoulder” and leaning forward 10cm. The organism can see four different objects: two natural objects (orange and plum) and two artefacts (glass and nail) (Fig. 1b). For simplicity, the image that is sent to the system is caused only by the objects and not the hand: this amounts to assuming that the hand is ignored on the basis of a suitable non-explicitly-simulated attention mechanism (Fig. 1c; cf. [36]). The simulated arm and hand have the same parameters of the iCub robot (<http://www.robotcub.org>). The model controls only 2-DOF of the hand: one for the thumb, whose DOF are controlled together proportionally to commands, and one for the four same-sized fingers, controlled as a whole “virtual finger” [11] again proportionally to commands. Reaching is not simulated as not relevant for the experiment (DOF of the arm are kept still).

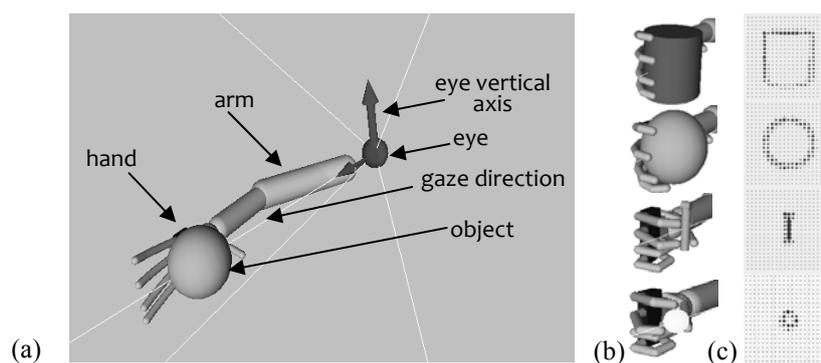


Fig. 1. (a) The simulated arm, hand, and eye interacting with a simulated object (orange). (b) Hand grips for four objects: glass, orange, nail, and plum. (c) The corresponding activation of PC neurons.

The activation of the output map of the model (premotor cortex) encodes (see Sect. 2.2) the desired hand's posture used to continuously feed the hand muscle models with "equilibrium points" [12]. Here, similarly to what is done in [13], single muscle models are simulated as simple Proportional Derivative controllers (PD) [14]. The equation of a PD muscle controller is as follows:

$$\mathbf{T} = \mathbf{K}_p \tilde{\mathbf{q}} - \mathbf{K}_d \dot{\mathbf{q}} \quad (1)$$

where \mathbf{T} is the vector of muscles' torques applied to the joints, \mathbf{K}_p is a diagonal matrix with elements equal to 300, $\tilde{\mathbf{q}}$ the difference vector between the desired joints' angular position and the current joints' angular position, \mathbf{K}_d is a diagonal matrix with elements equal to 10, and $\dot{\mathbf{q}}$ is the vector of current joints' angular speed. The PDs' action is integrated by a gravity compensation mechanism here implemented by simply ignoring the effects of gravity on the arm and hand.

2.2. Architecture and functioning of neural network model

The model is formed by nine 2D maps of 21×21 neurons each (Fig. 2). Visual cortex (V1) receives the visual pre-processed signal supplied by a simulated camera. Its neurons have an activation ranging in [0, 1] and encode the information about shape and colour of the foveated object obtained through three edge-detection Sobel filters [15]. Each filter is sensible to a particular component of the object's colour (red, green or blue: this simulates the functioning of the three kinds of cones in the human retina). The model assumes that the eye always foveates the target, in line with the current neuroscientific literature suggesting that primates tend to foveate the target objects with which they interact and that their brain exploits gaze centred reference frames as much as possible for sensorimotor coordination (see [16] for a review).

The neurons of parietal cortex (PC) encode the information about the object shape but not colour. To this purpose the neurons are activated with the average activation of the topologically correspondent RGB neurons of V1. This assumption is in accordance with recent neurophysiological data showing that the information about the object's shape plays a crucial role during learning and use of affordances related to objects [17][18].

The neurons of premotor cortex (PMC) encode the output of the system in terms of desired hand fingers' angles: these angles, mapped onto the 2 dimensions of the map, are "read out" as a weighted average of the neurons' position in the map with weights corresponding to the neurons' activation ("population code hypothesis", [19]). The PMC supports the selection of postures [20] on the basis of a dynamic competition between its leaky neurons

having lateral short-range excitatory connections and lateral long-range inhibitory connections [21]. When input signals from PC and PFC activate neurons of PMC, they tend to accumulate activation and form clusters (due to lateral excitatory connections) and, at the same time, to suppress other clusters (via lateral inhibitory connections). This dynamic process continues until a cluster succeeds in suppressing all other clusters, overcomes a threshold (set to 0.75), and so triggers the hand movement based on the reading out of the map:

$$s[j, t] = \overbrace{\sum_{i=1}^N w_{(PMC \rightarrow PMC)_{ji}} a_{PMC}[i, t]}^{Inner} + \overbrace{\sum_{i=1}^N w_{(PPC \rightarrow PMC)_{ji}} a_{PPC}[i, t]}^{Dorsal\ stream} + \overbrace{\sum_{i=1}^N w_{(PFC \rightarrow PMC)_{ji}} a_{PFC}[i, t]}^{Ventral\ stream} \quad (2)$$

$$u[j, t + \Delta t] = \left(1 - \frac{\Delta t}{\tau}\right) u[j, t] + \frac{\Delta t}{\tau} s[j, t] \quad a[j, t] = f[u[j, t]]$$

where $s[j, t]$, $u[j, t]$ and $a[j, t]$ are respectively the total signal, the activation potential, and the activation of neuron j at t time, Δt (set to 0.01s) is the integration time step (100 steps = 1s), τ (set to 0.3s) is the relaxation time, f is an activation function equal to $[\tanh[.]]^+$.

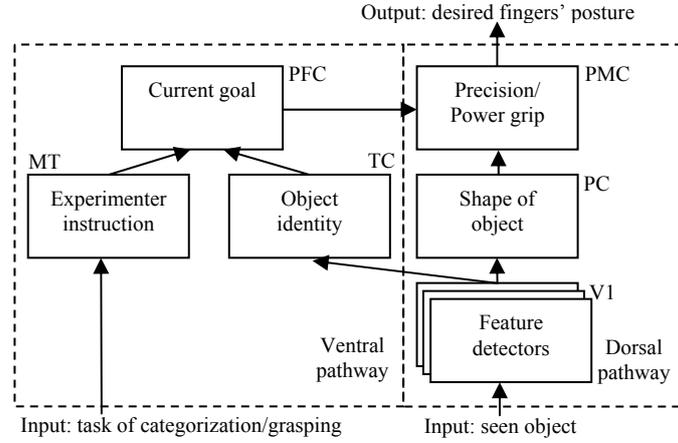


Fig. 2. Schema of the neural network model. V1 includes three RGB neural maps. Downstream V1, the model divides into two main neural pathways: the dorsal stream, implementing suitable sensorimotor transformations needed to perform action on the basis of perception, and the ventral stream, allowing a flexible control of behaviour due to the biasing effects of prefrontal cortex.

The *Inner* component of the formula accounts for signals received from lateral PMC connections with hardwired connection weights $w_{(PMC \rightarrow PMC)}$. These weights, excitatory for connections between neighbouring neurons and

inhibitory for connections between distant neurons, are set to fixed values on the basis of a Gaussian function and an inhibition term as follows:

$$w_{ji} = \exp\left[-\frac{(d[j, i])^2}{2\sigma^2}\right] - I \quad (3)$$

where w_{ij} is the weight between two neurons i and j of the map, $d[j, i]$ is the distance between the two neurons in the map “neural space” (where the measure unit is equal to 1 for two maximally-close neighbouring neurons), σ (set to 0.6) is the width of the Gaussian, and I (set to 0.9) is the inhibition term.

The *Dorsal stream* component accounts for the signals received from PC neurons modulated by the connection weights $w_{(PC \rightarrow PMC)}$; finally, the component *Ventral stream* accounts for the signals received from PFC neurons modulated by the connection weights $w_{(PFC \rightarrow PMC)}$. The reaction time is the time required by at least one neuron of the winner cluster of PMC to reach the threshold [21].

The neurons of temporal cortex (TC) encode objects’ identity. In accordance with visual physiology findings [22], from lower (V1) to higher levels (TC) of the visual hierarchy receptive field size and stimulus selectivity of neurons increase, whereas visual topography is progressively lost. In the model, TC is a Kohonen self-organising map (SOM) which activates as follows [23]:

$$a_j = \exp\left[-\frac{\|\mathbf{w}_{(V1 \rightarrow TC)j} - \mathbf{a}_{V1}\|^2}{2\sigma^2}\right] \quad (4)$$

where a_j is the activation of the TC neuron j , σ is the size of the clusters of active neurons equal to 0.55, $\mathbf{w}_{(V1 \rightarrow TC)j}$ is the vector of connection weights from V1 to TC neuron j , and \mathbf{a}_{V1} is the activation of V1 neurons. After learning, TC responds to different objects with different neuron clusters (see Sect. 2.3).

The neurons of medial temporal cortex (MT) encode the category of actions to be performed on objects, namely either those required by a grasping task, as performed by participants in everyday life, or those required by the psychological experiment specified by language. To this purpose, neurons of MT are activated with two random patterns with 20 neurons set equal to one and the rest equal to zero. The neurons of prefrontal cortex (PFC) encode information about the current goal of action depending on both the task (MT) and object identity (TC). To this purpose, PFC neurons are activated according to the Kohonen activation function of Eq. (4). The use of a Kohonen network for both TC and PFC is based on studies which suggest that these cortical areas are involved in high-level visual processing and categorization [24][26][35].

2.3. Learning phases

The organism undergoes two learning phases, one representing experience gathered in “normal life”, during which it learns to suitably grasp objects, and one representing the psychological experiment, during which the organism learns to trigger a power or precision grip on the basis of the objects’ category.

Before learning, the connection weights of the model are set to values uniformly drawn in $[0, 1]$. Learning during life involves learning of the affordance-based behaviour within the dorsal stream [27] and learning of objects’ identity in TC [28]. To this purpose, the four objects are repeatedly presented to the system in repeated trials during which MT is always activated with the pattern corresponding to the grasping task. Note that during this “life learning” the PFC and MT are activated, notwithstanding this would not be required to execute actions via the dorsal pathway, to avoid biasing the results when the ecological and experimental conditions are compared. In each trial the hand is open and the object is located in close to the hand palm, V1 performs Sobel-based colour-dependent edge detection of the object image, and PC performs colour-independent edge detection (that is it encodes object’s shape) by averaging the activation of RGB neurons of V1 with same topography.

The PC-PMC connection weights are developed using a Hebb covariance learning rule while the organism performs randomly-selected power/precision grip grasping actions in correspondence to the perceived object. This “motor babbling” [29][30] is a general learning process [37] for which the production of rather unstructured behaviours allow the formation of basic associations between sensory representations and motor representations [31]. Here, motor babbling is composed of these phases: (a) either a large (orange and glass both with a 7cm diameter) or small object (plum with a 2cm diameter or nail with a 8mm diameter) is set close to the system’s hand palm; (b) the hand moves its fingers around the object with constant torques (this is done by issuing suitable desired angles to the PD muscle models; objects are kept fixed in space to avoid that they slip away from fingers during closure); (c) when the fingers have been closed on the object (see Fig. 1b), the Hebb covariance learning rule reported below [32] is used to update the all-to-all connection weights between PC and PMC neurons so as to form associations between the object’s perceived shape (PC) and the corresponding hand posture (PMC):

$$\Delta w_{ji} = \eta (a_j - \bar{a}_j)(a_i - \bar{a}_i)(w_{max} - |w_{ji}|) \quad (5)$$

where η is a learning rate (set to 10), w_{max} (set to 0.2) keeps the connection weights within a small range, a_j is the activation of the PMC neuron j , a_i is the

activation of the PC neuron i , \bar{a}_i and \bar{a}_i are moving decaying averages of the neurons' activations calculated as $\bar{a}[t+\Delta t]=(1-\xi)\bar{a}[t]+\xi a$ (ξ is set to 0.8). This rule strengthens the connections between each couple of neurons which have both an activation above or both an activation below their own average activation, and weakens their connections in other cases. Within the ventral pathway, during motor babbling the V1-TC connection weights develop the capacity to categorise objects on the basis of a Kohonen learning rule [23][33]:

$$\Delta w_{ji} = \mu \Lambda[j, j^*] (a_i - w_{ji}) \quad \Lambda(j, j^*) = \exp\left[-\frac{d^2[j, j^*]}{2\sigma^2}\right] \quad (6)$$

where μ is a learning rate (set to 1), a_i is the activation of V1 neuron i , j is the index of an TC neuron, j^* is the index of the TC neuron with maximum activation (“winning neuron”), $\Lambda[j, j^*]$ is a proximity Gaussian function which determines the size of the cluster of neurons whose weights are updated, $d[j, j^*]$ is the Euclidean distance between j and j^* on the TC map, σ is the width of Gaussian function. During the model tests, the value of σ is set to a larger value for larger objects, in particular it is set within [0.5, 0.9] in proportion to the activation of V1 neurons for the various objects. This assumption is motivated by the following considerations. The Kohonen neural network is an approximation of the dynamic-field neural network of Eq. (2) and is used here because it is computationally faster and because it offers a well-understood learning algorithm. Contrary to the dynamic-field neural network, however, the Kohonen neural network has the implausible feature for which it forms clusters of active neurons having a *constant* number of units and overall activation level. As the total activation of clusters may have important effects on RTs, and this was important for the goals of the paper, this limit of the Kohonen network is overcome with the assumption on the variable σ . Note that a similar assumption is done also for PFC (see below). Having a larger number of active neurons in correspondence to larger objects seems a better approximation of what might happen in real brains (e.g., in this way information on objects size is encoded in terms of overall activation of neurons, cf. Hu and Goodale, 2000).

During the psychological experiment learning involves acquiring suitable “goal representations” in PFC, that is representations of which action to select (stored in the PFC-PMC connections) in correspondence to which combination of task and object identity currently tackled by the organism (stored in the (MT, TC)-PFC connections). To this purpose, the four objects are repeatedly presented to the system in multiple trials during which (a) MT is always activated with the pattern corresponding to the categorisation task, and (b) the hand has performed the grip requested by the psychological experiment. The connection weights between (MT, TC)-PFC are updated using the modified Kohonen algorithm of Eq. (6). Similarly to TC, also within PFC the use of the

modified Kohonen algorithm allows obtaining larger clusters of activated neurons for larger objects. In this way the model assumes that the ventral stream stores information about the object size in terms of number of activated neurons (cf. [34]). During the objects presentation, accompanied by the hand closure requested by the categorisation task of the simulated psychological experiment, also the connection weights between PFC and PMC are updated, in this case on the basis of the Hebb covariance rule of Eq. (5). This allows the system to associate the particular combination of task (MT) and object identity (TC) with the suitable action required to correctly categorise the observed object (PMC).

3. Results

The model reproduces the experimental results of [1] (Fig. 3). An ANOVA on response times was performed with two factors: congruency (congruent vs. incongruent) and object size (large vs. small). Participants were ten different simulated organisms trained and tested with ten different random-number generator seeds. In agreement with the experiments run with real subjects, both factors were statistically significant: RTs were faster in congruent than in incongruent trials ($p < 0.01$) and for large than for small objects ($p < 0.01$).

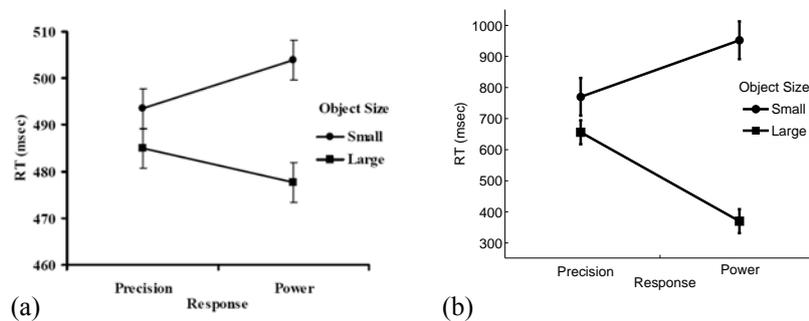


Fig. 3. Reaction times (y-axis) versus kind of grip (x-axis). (a) Real experiment [1] (copyright of Taylor & Francis, Visual Cognition, <http://www.informaworld.com>). (b) Simulated experiment.

The analysis of the system during “life” (i.e. while performing a grasping task) shows that the dorsal pathway tends to trigger an action on the basis of the affordances elicited by it (a power grip for a large object, a precision grip for a small object). In particular, after the learning phase, when organisms see an object the neurons of V1 become active by encoding shape and colour, and thus the neurons of PC become active by encoding the shape. The activation of PC causes a neuron cluster in PMC to gain activation until it reaches the action-

triggering threshold. This leads to the execution of either a power or precision grip through the muscle-models and the simulated hand. In parallel, when the task requires grasping objects the ventral pathway evokes the same congruent action in PMC as the dorsal pathways (see Fig. 4a). In particular, TC activates four different clusters of neurons, one for each different object, PFC does the same, and each PFC cluster evokes the action congruent with the object.

In the case of the categorisation task, TC activates four different clusters of neurons, one for each different object, as during “life”. On the contrary, PFC activates four different clusters with respect to what happens during “life”, as now the MT activation pattern indicates to the system that it needs to respond to each object with an action which depends on its category (artefact vs. natural) and not on its size (small versus large). As a consequence, in incongruent trials the ventral pathway evokes an action different with respect to the dorsal pathway (e.g. a precision grip to categorise as artefact an orange) via a suitable bias issued by the PFC to the PMC, thus causing a conflict within the latter (Fig. 5b). As the PFC-PMC signal is stronger than the PC-PMC signal, the bias from PFC wins the competition (e.g. by triggering a precision grip) but the RTs is longer with respect to the congruent cases. In fact, when PFC and PC signal clusters mismatch they lead to a slower charge of the PMC leaky neurons, and hence reaching of the threshold requires more time than when they match.

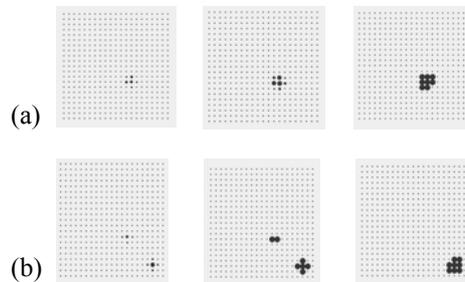


Fig. 4. (a) Activation of PMC in a congruent trial. (b) Activation of PMC in an incongruent trial.

The results also show faster RTs for the larger objects, as in real experiments. The reason is that large objects activate more neurons of V1 than small ones, and these activate a larger number of neurons in PC, and, in turn, in PMC. The signal arriving to PMC via the ventral pathway is also greater for large objects than for small ones due to the use of the modified Kohonen learning rule. Both phenomena tend to produce a faster “charge” of the leaky neurons of the PMC and hence faster RTs (see Eq. (1)). In incongruent trials, the two signals from PFC and PC to PMC do not overlap but are both large: a

competition between two large activation patterns again leads to faster RTs in PMC than when there is a competition between two small patterns. However, these RTs are slower than RTs of congruent trials.

4. Conclusion

The proposed model successfully reproduces the experimental results of [1], in particular faster reaction times in congruent conditions vs. incongruent ones, and with large objects vs. small objects. The architecture of the model *integrates* three important principles, namely the interplay between ventral and dorsal visual pathways, the top-down biasing effect of prefrontal cortex on selection of actions, and the dynamic-field competitive processes, which make it quite general and suitable to tackle other empirical experiments in future work.

Acknowledgments

This research was supported by the EU FP7 Project ROSSI, contract no. 216125-STREP.

References

- 1 M. Tucker and R. Ellis, *Vis Cogn* **8**, 769 (2001).
- 2 L. W. Barsalou, *Annu Rev Psychol* **59**, 617 (2008).
- 3 G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, M. Matelli, *Exp Brain Res* **71**, 491 (1988).
- 4 Berthoz, *Philos T R Soc B* **352**, 1437 (1997).
- 5 J. Gibson, *The ecological approach to visual perception* (1979).
- 6 M. Borghi, C. Bonfiglioli, L. Lugli, P. Ricciardelli, S. Rubichi and R. Nicoletti, *Neurosci Lett* **411**, 17 (2007).
- 7 U. Castiello, *Trends Cogn Sci* **3**, 264 (1999).
- 8 E. K. Miller and J. D. Cohen, *Annu Rev Neurosci* **24**, 167 (2001).
- 9 P. S. Churchland and T. J. Sejnowski, *The Computational Brain* (1993).
- 10 J. Grèzes, M. Tucker, J. Armony, R. Ellis and R. E. Passingham, *Eur J Neurosci* **17**, 2735 (2003).
- 11 M. A. Arbib, Grounding the mirror system hypothesis for the evolution of the language-ready brain, in *Simulating the Evolution of Language*, eds. A. Cangelosi and D. Parisi, 229 (2002).
- 12 G. Feldman, *J Mot Behav* **18**, 17 (1986).
- 13 N. E. Berthier, M. T. Rosenstein and A. G. Barto, *Psychol Rev* **112**, 329 (2005).
- 14 D. Bullock and S. Grossberg, Vite and Flete: neural modules for trajectory formation and postural control, in *Volitional Action*, ed. W. Hershberger (1989).

- 15 I. Sobel and G. Feldman, A 3x3 isotropic gradient operator for image processing, *Presentation for Stanford Artificial Project* (1968).
- 16 R. Shadmehr and S. P. Wise, *The Computational Neurobiology of Reaching and Pointing* (2005).
- 17 Murata, V. Gallese, G. Luppino, M. Kaseda and H. Sakata, *J Neurophysiol* **83**, 2580 (2000).
- 18 R. T. Oliver, E. J. Geiger, B. C. Lewandowski and S. Thompson-Schill, *J Vision* **5**, 610 (2005).
- 19 A. Pouget, P. Dayan and R. Zemel, *Nat Rev Neurosci* **1**, 125 (2000).
- 20 M. L. Platt, *Nature* **400**, 233 (1999).
- 21 W. Erlhagen and G. Schooner, *Psychol Rev* **109**, 545 (2002).
- 22 Tanaka, *Annu Rev Neurosci*, **19**, 109 (1996).
- 23 T. Kohonen, *Self-Organizing Maps* (1997).
- 24 E. K. Miller, D. J. Freedman and J. D. Wallis, *Philos T R Soc B*, **357**, 1123 (2002).
- 25 D. Milner and M. A. Goodale, *Neuropsychologia* **46**, 774 (2008).
- 26 D. J. Freedman, M. Riesenhuber, T. Poggio and E. K. Miller, *J Neurosci*, **23**, 12, 5235 (2003).
- 27 P. Cisek, *Philos T R Soc B* **362**, 1585 (2007).
- 28 H. R. Rodman, *Cereb Cortex* **4**, 484 (1994).
- 29 D. Caligiore, D. Parisi and G. Baldassarre, Toward an integrated biomimetic model of reaching, in *Proceedings of 6th IEEE International Conference on Development and Learning*, eds. Y. Demiris, B. Scassellati and D. Mareschal (2007).
- 30 C. von Hofsten, *Dev Psychol*, **18** (1982).
- 31 Piaget. *The origins of intelligence in children* (1952).
- 32 T. J. Sejnowski, *J Math Biol* **4**, 303 (1977).
- 33 H. Ritter, T. Martinetz and K. Schulten, *Neural Computation and Self-Organizing Maps - An Introduction* (1992).
- 34 Y. Hu and M. A. Goodale, *J Cognitive Neurosci* **12**, 856 (2000).
- 35 K. Shima, M. Isoda, H. Mushiake and J. Tanji, *Nature* **445**, 315 (2007).
- 36 D. Ognibene, C. Balkenius and G. Baldassarre, Integrating epistemic action (active vision) and pragmatic action (reaching): a neural architecture for camera-arm robots, in *Proceedings of the Tenth International Conference on the Simulation of Adaptive Behavior (SAB2008)*, eds. M. Asada, J.C.T. Hallam, J.-A. Meyer and J. Tani (2008).
- 37 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, in *International Conference on Cognitive Systems*, eds. R. Dillmann, C. Maloney, G. Sandini, T. Asfour, G. Cheng, G. Metta, A. Ude (2008).