

How active perception and attractor dynamics shape perceptual categorization: a computational model

Nicola Catenacci Volpi, Jean-Charles Quinton, Giovanni Pezzulo

▶ **To cite this version:**

Nicola Catenacci Volpi, Jean-Charles Quinton, Giovanni Pezzulo. How active perception and attractor dynamics shape perceptual categorization: a computational model. Neural Networks, 2014, 60, pp.1-16. 10.1016/j.neunet.2014.06.008 . hal-01966566

HAL Id: hal-01966566
https://hal.science/hal-01966566

Submitted on 28 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

How active perception and attractor dynamics shape perceptual categorization: a computational model

Nicola Catenacci Volpi[a,*], Jean Charles Quinton[b,c], Giovanni Pezzulo[d]

[a] *IMT Institute for Advanced Studies, Piazza S. Ponziano 6 - 55100 Lucca, Italy*
[b] *Clermont University, Blaise Pascal University, Pascal Institute, BP 10448, F-63000 Clermont-Ferrand, France*
[c] *CNRS, UMR 6602, Pascal Institute, F-63171 Aubiere, France*
[d] *Istituto di Scienze e Tecnologie della Cognizione - CNR, Via S. Martino della Battaglia, 44 - 00185 Rome, Italy*

Abstract

We propose a computational model of perceptual categorization that fuses elements of grounded and sensorimotor theories of cognition with dynamic models of decision-making. We assume that category information consists in anticipated patterns of agent-environment interactions that can be elicited through overt or covert (simulated) eye movements, object manipulation, etc. This information is firstly encoded when category information is acquired, and then re-enacted during perceptual categorization. The perceptual categorization consists in a dynamic competition between attractors that encode the sensorimotor patterns typical of each category; action prediction success counts as "evidence" for a given category and contributes to falling into the corresponding attractor. The evidence accumulation process is guided by an active perception loop, and the active exploration of objects (e.g., visual exploration) aims at eliciting expected sensorimotor patterns that count as evidence for the object category. We present a computational model incorporating these elements and describing action prediction, active perception, and attractor dynamics as key elements of perceptual categorizations. We test the model in three simulated perceptual categorization tasks, and we discuss its relevance for grounded and sensorimotor theories of cognition.

Keywords: Hopfield networks, perceptual categorization, prediction, active vision, dynamic choice

1. Introduction

Cognitive scientists and neuroscientists have widely studied how the brain categorizes and recognizes objects and entities. Traditional cognitive psychology theories propose that categories are stored in the form of sets of rules that define the category (Trabasso and Bower, 1968), category prototypes that average across category elements (Rosch, 1975), sets of exemplars that correspond to specific category elements (Medin and Schaffer, 1978), or a combination of all them.

*Corresponding author. Address: Via S. Martino della Battaglia, 44 - 00185 Rome, Italy. Phone Number: 0039 06 44596208

Email addresses: nicola.catenacci@imtlucca.it (Nicola Catenacci Volpi), j-charles.quinton@univ-bpclermont.fr (Jean Charles Quinton), giovanni.pezzulo@cnr.it (Giovanni Pezzulo)

More recently, grounded theories of cognition have put categorization research into a broader perspective, arguing how the abilities of perceiving, categorizing, and thinking about objects and events are highly related. According to this view, perceptual and motor processes (rather than amodal symbols) constitute the conceptual content of categories and concepts, including abstract ones. Once acquired through agent-environment interaction, this conceptual content can be re-enacted to support off-line thinking and cognition, determining so-called *situated simulations* (Barsalou, 1999, 2008; Pezzulo et al., 2011, 2013).

Sensorimotor theories stem from similar assumptions but further emphasize the importance of action dynamics. They assume that actions are constitutive of perception and categorization, and describe object perception in terms of interaction dynamics and stable patterns of actions and sensory stimulations, or *sensorimotor contingencies* (O'Regan and Noe, 2001). For example, the sight of a glass produces a coherent pattern of sensory stimulations (as an effect of the eye's actions), and the mastery of such contingencies corresponds to the knowledge of what is a glass. Support from this view comes from experiments showing the importance of action (e.g., eye movement) dynamics in shaping the categorization process (Ballard, 1991; Hayhoe and Ballard, 2005; Rothkopf et al., 2007; Yarbus, 1967).

In a similar vein, action-based approaches emphasize that knowledge of the external world consists in sets of "dispositions to act" as produced by action-outcome mechanisms, and that object knowledge consists in the anticipated patterns of actions and perceptions produced by an interaction with them (Bickhard, 1993; Grush, 2004; Pezzulo, 2008, 2011). For instance, a sponge can be understood in terms of a characteristic (sequence of) action-outcome relation, such as the anticipated softness one expects when squeezing it. These action-outcome relations have been linked to the concepts of internal forward models (Desmurget and Grafton, 2000; Kawato, 1999) and ideomotor codes (Hommel et al., 2001); see also (Maye and Engel, 2011; Pezzulo and Calvi, 2011; Roy, 2005). Numerous studies support the idea that the same action-outcome links adopted in the on-line interactions can also be reused off-line to *mentally simulate* an interaction, essentially recruiting the same brain mechanisms for *motor cognition* (Jeannerod, 2001, 2006).

In this article we offer a theory of perceptual categorization that distills key concepts of grounded, sensorimotor and action-based theories of cognition and integrates them with dynamic and competitive models of choice. We propose that categories are coded in terms of the associated action-outcome sequences, not in purely sensorial terms. Specifically, an object category is linked to (predictable) sequences of saccades, grasp movements, or a combination of them. This sensorimotor information is firstly acquired during situated agent-object interactions and can be successively re-enacted to guide perceptual processing and categorization, in a process that resembles the sampling of environmental information under the guidance of categorical hypotheses (Barsalou, 1999; Pezzulo et al., 2013).

In this view, agent-object relations can be described as sequences of actions and resulting sensations, or action-outcome pairs. In keeping with grounded cognition theories, we assume that this information is acquired when the agent interacts with exemplars of the category (Barsalou, 1999). For instance, during interactions with a sponge an agent learns action-outcome relations: how a sponge feels when it is squashed, how it looks if it is foveated to the left or right, etc. Once learned, the same sensorimotor processes used to explore (e.g., visually or haptically) and interact with objects also realize the object categorization process; for instance, a sponge is recognized when the agent successfully reuses the stored action-outcome relations associated to earlier sponge uses. The same information can be reused to mentally simulate interactions with the same objects in their absence (Pezzulo, 2011). Action-outcome relations are maintained in the internal models used to interact with objects, in a modal format; more frequently, objects link to multimodal information acquired using different effectors (e.g., eye and hand). Categorization profits from both overt exploration (e.g., physical manipulation of a sponge) and mental simulation (e.g., just

anticipating the interaction), which according to grounded theories of cognition recruit the same brain processes.

The pragmatic view of categorization that we propose emphasizes the importance of previous interactions, like exemplar and prototype theories of categorization. At the same time, it reverses the perception-categorization-action pipeline of traditional cognitive theories, and proposes that action is part and parcel of perception and categorization rather than being successive to the categorization. In sum, our approach assumes that action-outcome representations are constitutive of the conceptual content of categories, at least for categories that can be readily mapped to possible interactions.

1.1. The mechanics of situated categorization

Up to now we have introduced our proposed theory of categorization by referring abstractly to action-outcome patterns. Now we discuss how this information is elicited during situated interactions with the to-be-categorized object and how it influences the moment-by-moment dynamics of the categorization.

There is ample consensus that perceptual decision-making and categorization are dynamic and competitive processes in which evidence is accumulated in favor or against the competing alternatives (e.g., deciding if a visual stimulus is a cat or a dog). The widely adopted drift-diffusion model describes choice as a competitive process of accumulation of evidence up to a criterion; when the criterion is reached, action can start (Ratcliff, 1978; Ratcliff and Rouder, 1998); see also (Bogacz et al., 2006; Usher and McClelland, 2001; Wang, 2002) for descriptions of plausible neural implementations of diffusion-to-bound and related mechanisms. Several models of perceptual categorization invoke the same dynamic mechanisms but differ on what they consider to be the relevant dimensions along which evidence is accumulated. An influential model (Nosofsky and Palmeri, 1997) describes perceptual categorization as a dynamic competition between exemplars, linking to the exemplar models of categories described earlier. Another model (Lamberts, 2000) uses the same principles of dynamic accumulation of evidence, but focuses on competition between stimuli features rather than exemplars.

Diffusion-to-bound models have been extremely successful in explaining behavioral data and map nicely to the brain substrate (Gold and Shadlen, 2001, 2007). This leads to the idea that core mechanisms of decision-making (based on evidence accumulation) could have been preserved to support increasingly more complex and abstract decisions and categorizations (Cisek, 2012; Shadlen et al., 2008). However, they largely abstract from the way evidence is accumulated. They often point to a bottom-up process in which a stimulus is repeatedly probed to obtain multiple samples and do not model active perception dynamics or overt exploration (but see Krajbich et al. (2010)). Our proposed model extends these theories and describes categorization as a dynamic and competitive process that builds on evidence elicited through active perception.

Within the dynamic and competitive categorization process that we discussed, it is often assumed that evidence accumulation follows a sequential sampling rule (Ratcliff, 1978), which corresponds to an optimal statistical test. We assume that active perception dynamics bias the evidence accumulation process; this process is not random but rather it recapitulates the agent-object interactions that firstly created the agent's categorical concepts. In other words, the visual exploration of an object consists in an attempt to re-create and re-elicit the same action-perception patterns that were established when an object category was acquired, and the elicitation of the same (predicted) stimuli counts as evidence for the category. The proposed model (sketched in Figure 1) is based on three main assumptions that we discuss below.

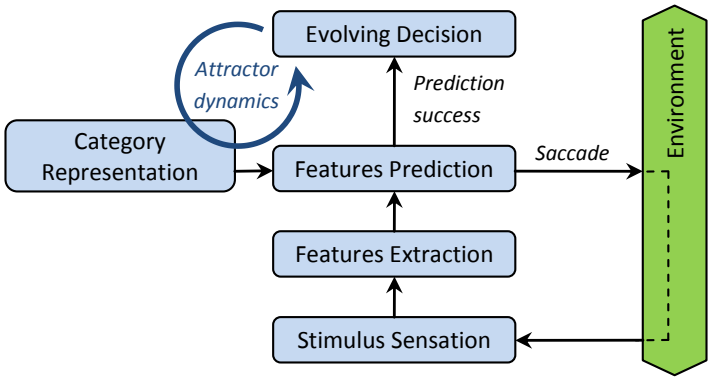


Figure 1: Proposed dynamic model of perceptual categorization. See main text for explanations.

1.1.1. Action dynamics shapes the ongoing categorization

A first aspect that distinguishes our model from previous ones is that it gives motoric and active perception processes a key role in the categorization and decision-making process. The theories of decision-making that we have considered so far describe evidence accumulation as a bottom-up process segregated from action dynamics; in this two-stage approach, actions are executed only when decisions are completed. However, recent experiments support a *continuous flow model* (Coles et al., 1985), in which evidence-accumulation and the preparation of corresponding actions are not segregated, but partial decisions steer actions, so that movement trajectories can be informative of the underlying decision dynamics and uncertainties (Barca and Pezzulo, 2012; Resulaj et al., 2009; Song and Nakayama, 2009; Spivey, 2007)[1].

If the flow was unidirectional, from decision to motor processes, then it could be regarded as a nuance without importance for the choice itself. However, there are reasons to believe that action dynamics feed back and influence perceptual processes and choice, and that processes of motor preparation, planning, and execution can influence the (perceptual) decision-making in many ways. The *premotor theory of attention* (Rizzolatti et al., 1987) suggests that motor processes have modulatory effects on sensory processing; for instance, planned actions can direct attention by priming relevant stimulus dimensions (Fagioli et al., 2007). Furthermore, theories of *active vision* (Ballard, 1991) emphasize that the perceptual flow is not passive but actively selected, and overt perceptual processes (e.g. gaze allocation) bias perception (and indirectly the accumulation of information for perceptual decision-making). Finally, recent evidence indicates that it is possible to influence object perception and categorization by influencing overt attention, suggesting that action dynamics (e.g., eye movements) are constitutive elements of perception (Kietzmann et al., 2011), as also proposed by sensorimotor contingencies theory (O'Regan and Noe, 2001). All these elements point to the conclusion that action dynamics (and specifically eye movements in our model, see figure 1) affect the evolving decision.

The idea that categorization is a situated activity based on active perception and agent-environment interactions is present in several models using evolutionary robotics techniques

[1]The continuous flow produces adaptive advantages by linking perceptual processing to motor preparation and execution, making an organism more responsive to real-world challenges. This idea seems plausible if one considers that the perceptual system of living organisms evolved to support rapid action selection in dangerous environments, not for discriminating between stimulus categories per se (Cisek and Kalaska, 2010; Pezzulo and Castelfranchi, 2009).

(Gigliotta and Nolfi, 2008; Mirolli et al., 2010; Nolfi and Marocco, 2002; Tuci et al., 2010). Furthermore, the importance of situated interactions for categorization has been recognized in dynamical systems (Quinton et al., 2013; Schoener, 2008; Spivey, 2007; Strauss and Heinke, 2012; Tipper et al., 2000) and probabilistic models (Maye and Engel, 2011). Our model incorporates aspects of these proposals while retaining key aspects of the evidence-accumulation framework discussed earlier.

1.1.2. Evidence accumulation and action guidance are both based on prediction success

A second aspect that distinguishes our approach from previous categorization models is that it considers predictive processes central in the choice. Predictive coding models have long assumed the importance of predictive dynamics for perception and categorization, using prediction success for an error minimization process (Friston, 2005; Rao and Ballard, 1999). In our model prediction dynamics are central, too, for two main reasons.

First, we assume that the relevant dimensions over which evidence is accumulated is *prediction success* of the action-outcomes mechanisms associated with each object category. To understand how this is possible, consider the links between "category representation" and "features prediction" and between "features prediction" and "evolving decision" in Figure 1. During a decision-making task, the overt (e.g., visual or haptic) exploration or even its simulation (e.g., when one simulates grasping an object) produces anticipated patterns of motoric and sensory stimulations that are plausibly acquired during category learning, such as the prediction that if one moves the eyes to the right, it will see a given feature, if one then moves again to the right, it will see another feature, and so on (see the link between "category representation" and "features prediction" in Figure 1). In turn, as the elicited action-outcome pairs constitute an embodied form of knowledge of the categories or objects, the predictions they produce are constitutive for the choice (see the link between "features prediction" and "evolving decision"). In other words, the recognition of an object (e.g. a straight line) boils down to the accumulation of a series of successful predictions (e.g., expecting to find the same features when saccading in a fixed given direction, see O'Regan and Noe, 2001). Similarly, the discrimination of a sponge vs. a brick depends on the prediction success of action-outcome mechanisms that predict squeezing or not squeezing following a (real or simulated) grasping action. This mechanism is similar to the prediction-error minimization of predictive coding theories (Friston, 2005; Rao and Ballard, 1999) but uses action outcome prediction dynamics within an overall evidence accumulation process. The idea of using predictive power of the features to drive categorization is not novel (Kruschke, 1992); what is novel in our model is that, in keeping with action-based and sensorimotor theories, the predictions are generated by action-outcome mechanisms and elicited through overt or covert exploration. In other words, we assume that the same predictive processes implied in performing object-directed actions are reused to recognize the same object (or even to recognize actions performed by others (Dindo et al., 2011; Wolpert et al., 2003)).

Second, we assume that predictive information drives overt stimulus exploration (e.g., eye movements). Studies of eye movements during concept formation (Nelson and Cottrell, 2007; Rehder and Hoffman, 2005) and categorization (Nelson et al., 2010) reveal that knowledge of anticipated perceptual patterns is key to guiding overt exploration towards the most informative dimensions. The process of evidence accumulation can thus be guided by processes that tend to test hypotheses and predictions and minimize uncertainty relative to the choice alternatives rather than selecting (only) the most salient stimuli features (Friston et al., 2012; Geisler, 2011). In Figure 1 this process is represented by the link between "features prediction" and "stimulus sensation".

1.1.3. *Evidence accumulation follows attractor dynamics that mimics structured agent-environment interactions*

The last aspect that characterizes our approach is the nature of the evidence accumulation process. As we have discussed, evidence accumulation is a process of re-creation and re-enactment of the interactive knowledge acquired when interacting with objects of a given category. It is important to notice that interactions with objects give rise to highly structured sequences of actions and resulting sensations (not just random samplings). The same structure is plausibly incorporated in the neuronal patterns encoding category information, and can be reused during evidence accumulation.

While neuronal mechanisms can encode structured information in multiple ways, several studies suggest that attractor dynamics could have a key role (Bassett and Gazzaniga, 2011; Churchland et al., 2010). For example, in a monkey classification task using familiar images and morphed stimuli, the authors report that the conversion of graded visual information into a category can be mathematically characterized in terms of local attractor dynamics within populations of inferior temporal neurons (Akrami et al., 2009). This evidence is relevant for our study as the task we model is very similar (see below).

Attractor models have long been recognized as relevant as brain models for categorization (Amit et al., 1997; Miyashita and Chang, 1988; Sakai and Miyashita, 1991) and are widely used to model perceptual decisions in cognitive psychology (Spivey, 2007) and neuroscience (Wang, 2002). In these models, the way evidence is accumulated reflects the attractor dynamics of (e.g., the "structure" of the information encoded in) the neuronal populations responsible for the choice. In our proposed model, attractor dynamics govern the way category representation, features prediction and the evolving decision are linked (see the circular arrows in Figure 1).

1.2. Structure of the article

In the rest of the article, we incorporate the model presented so far in a computational architecture (see Section 2) and test it in three tasks (see Section 3). The first task is a visual categorization experiment consisting in recognizing if a stimulus (normal or morphed) is a dog, giraffe, horse, or cat. This task helps illustrating the main characteristics of the model, and shows how the decision can be conceptualized in terms of attractor dynamics steered by action-outcome predictions. In the second task, to support our claims on the importance of prediction and action dynamics, we show that categorization of ambiguous figures is significantly affected by the specific sequence of actions (fixations) used during the task, similarly to the study reported by (Kietzmann et al., 2011). In the third task we show how the model can qualitatively reproduce the data reported in (Akrami et al., 2009) on the categorization of stimuli having various levels of ambiguity. Finally, in Section 4 we draw our conclusions.

2. Method: The Computational Model

The proposed computational architecture essentially connects several (category-selective) sensorimotor feature predictors through a Hopfield network (Hopfield, 1982; Amit, 1989), and performs a categorization by accumulating evidence on predictors' success during the visual exploration of the to-be-categorized object. In the scenario we address in this article, the categorization task consists in determining whether a stick figure (input image) represents a giraffe, a horse, a cat or a dog; features represent picture parts.

The model components and their interactions are sketched in Fig. 2.

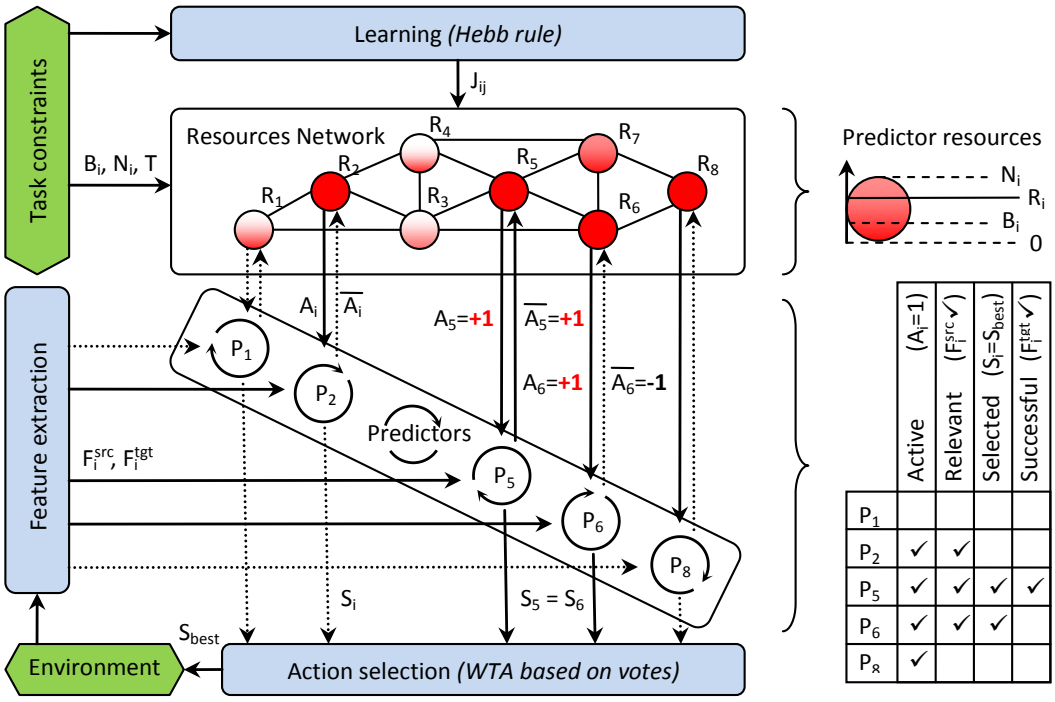


Figure 2: Attractor Predictors Network. After the learning phase which determines the weights J_{ij} within the resources network, the network modulates the activity of the predictors, themselves interacting with the environment. Resources R_i above threshold N_i (*fully red nodes*) leads to the activation of the associated predictor P_i. If a predictor is moreover relevant (source feature F_i^{src} found), if its action gets selected ($S_{best} = S_i$) and is then confirmed (target feature F_i^{tgt} is found), it gives a positive feedback to the network (prediction success $\overline{A_i} = +1$). The 5 possible predictor configurations are represented on the figure, with *plain lines* representing activation/confirmation signals. The reason why here both P5 and P6 are selected is that predictors can share the same starting feature and vote for the same action having different target features (see the S5=S6 equation within the figure). P_5 and P_6 are active, relevant and see their action selected, but then only P_5 has its prediction confirmed. See main text for explanation.

2.1. Motor outputs and sensory inputs to the architecture

Adopting an active vision approach to categorization, the architecture is only given access to the stimuli through a reduced field of view. Because of the size of this area relative to stimuli, the system needs to explore stimuli through saccades in order to obtain information and discriminate between them. This sensory limitation is however compensated by active capabilities. The computational architecture has control over this small foveated area, that can be freely moved around the stimulus to categorize. Actions thus consist in saccades (S) that instantaneously move the foveated area from one position to another. As a consequence, the proposed architecture not only needs to exploit salient information acquired through bottom-up processes (at each fixation point) (Itti et al., 1998), but must also direct top-down attentional processes, by selecting relevant features in the visual field (covert) and controlling eye movements (overt). Other computational architectures use such a combination to efficiently extract information from the environment, but the selection and sequence of saccades is generally segregated from the classification of stimuli (Frintrop and Jensfelt, 2008).

The visual input from the foveated area is preprocessed. This is first to compensate for the lack of the kind of highly robust visual system and life-long learning found in humans. It keeps

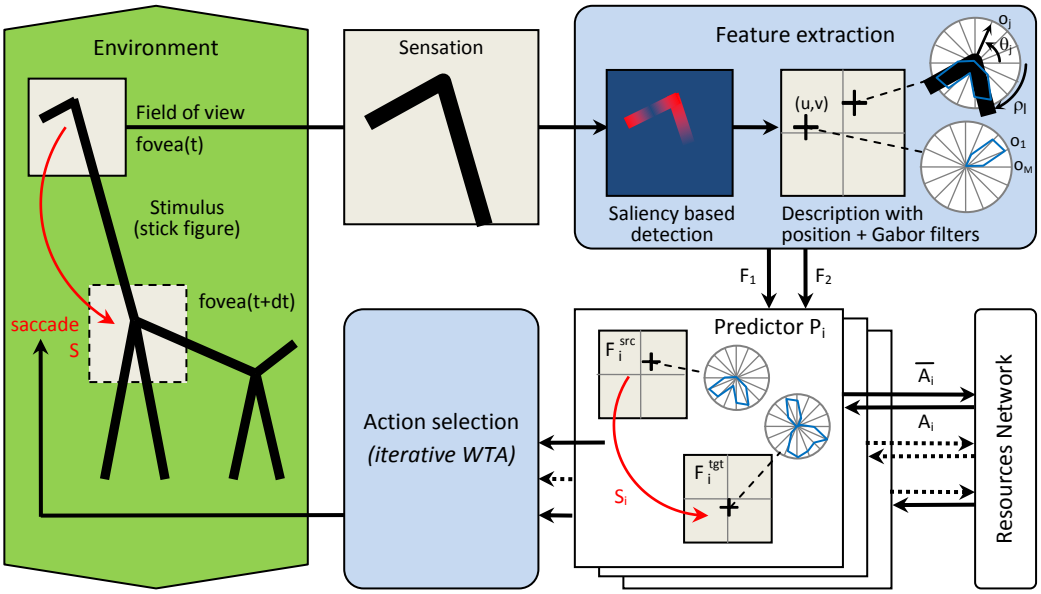


Figure 3: Representation of the feature extraction process and of the predictors coding (*subpart of Fig. 1*). At any time t, a variable number of features lying within the foveated area of the visual environment (here F_1 and F_2) are detected and described by their position relative to the center of the fovea (u, v) (*crosses*), and the activity of M fixed orientation detectors (o_j) (*polar diagram*). If active, a predictor P_i can then compare these observed features with its source feature (F_i^{src}) (*context*) and propose a saccade S_i that may be selected through a winner-take-all competition, thus shifting the fovea to another position at time $t + dt$, so it can test the presence of the expected target feature (F_i^{tgt}).

the system complexity high enough for the phenomena we are interested in to emerge, yet allows for an intuitive understanding of the overall system dynamics and predictors role. Finally and for a more technical reason, this allows real time simulation of the model on standard computer architectures.

As we focus on categorization and decision-making processes, we provide the system with the most informative features available on the specific class of stimuli considered (see Fig. 3 for an example). Neuro-inspired saliency maps were computed over stick figure images. By saliency, we here refer to a human inspired vision algorithm where multi-scale luminance contrasts, orientations and color contrasts (not relevant in our case) are extracted from the visual input, and later combined in a single conspicuity map (Itti et al., 1998). A slightly altered version of such algorithm is used here, where both on-off and off-on intensity contrast detectors are considered in order to deal with black figures on a white background. The fact that the same processing must occur at different scales, orientations and positions of the visual field are common place in cortical areas (e.g. MST for motion perception (Clifford et al., 1999)). This approach uses population coding and pooling, where a discrete set of orientations and scales are sufficient to encode for continuous changes in the sensory flow. Distributed systems can then rely on their dynamics to interpolate between different scales, so that resulting detectors and descriptors reflect continuous variations (Quinton and Girau, 2012).

As expected and due to the shape, structure and black-and-white nature of the stimuli used in this paper, oriented Gabor filters on a single scale had the strongest response, and saliency is concentrated around the joints of the stick figures. In order to lower the dimensionality of the

input, only a set of feature points (F_i) corresponding to the visible joints of the stick figure are retained for each fixation. Each feature is described by its coordinates within the retinal image (u, v), as well as a vector synthesizing the oriented Gabor filter responses away from the joint $(o_j)_{j\in[1,M]}$ (with M fixed to 16 in the experiments). In practice, Gaussian tuning curves with wide selectivity profiles along orientations have been used to directly convert an arbitrary set of orientations into a fixed number of correlated activities from the stick figure description (see Eq. 1 and Fig. 3). Even with these highly selective features, the non triviality of the discrimination process must once again be underlined. There is for instance no way for the system to easily distinguish between front and back legs without saccading across the stimulus.

$$o_j = \max_l \exp - \left(\frac{M(\theta_j - \rho_l)}{2\pi} \right)^2 \tag{1}$$

where $\theta_j = -\pi + j\pi/M$, and ρ_l is the set of angles formed by the sticks starting from the considered joint.

Features are thus fully described by $F_i = (u, v, o_1, \ldots, o_M)$ and two features (F_1, F_2) can be compared according to a similarity measure defined as:

$$\sigma(F_1, F_2) = 1 - e^{-\frac{\|F_2 - F_1\|^2}{\gamma^2}} \tag{2}$$

where $\|.\|$ is a norm in $\mathbb{R}^{M+2}$ with an adequate weighting of the various dimensions involved. This weighting must simply ensure that an improved precision for the orientation filters (increase of M) will not make their influence dominant over the position discrepancy within the visual field (dimensions u and v). This is implemented using an exponentially weighted sum to make the topology smoother, with weight 0.2 for the coordinates (u, v) in the experiments and 1.0 for the orientations $(o_j)_{j\in[1,M]}$ (both based on an estimation of the variability of the components values). Moreover, γ is chosen to be proportional to the number of the dimensions.

Please note that we could have used other well known descriptors from computer vision such as SIFT, which is also based on multiscale processing (Lowe, 1999). The distributed cognitive model developed in this paper could however directly benefit from saliency based feature extraction. Indeed, and although we here focus on a single scale in order to underline the specificities of the cognitive model, the similarity σ defined in Eq. 2 could be easily extended and the number of predictors multiplied to make the recognition scale-invariant. Finally, it is also possible to dynamically tune such preprocessing to make it more reactive to specific features, a capability particularly well suited to active and predictive perception systems. Gain neurons (Salinas and Sejnowski, 2001) or a direct modulation of saliency maps (Frintrop and Jensfelt, 2008) can be used for such purpose.

2.2. Predictors

Building on the visual features provided to the system and the potential actions that can be performed, local predictors can be defined. Predictors represent particular sensorimotor regularities expected while interacting with the to-be-categorized object. In other words, predictors are specialized units able to implement modality-specific sensorimotor strategies. A predictor P is here defined by a triple (F^{src}, S, F^{tgt}) where F^{src} is the source feature, S is the motor command and F^{tgt} is the expected target feature. When applied to active vision, each predictor (P_i) suggests a saccadic motor command (S_i) that moves the fovea from an area where the feature F_i^{src} was visible to an area where the predictor expects to find the feature F_i^{tgt} (see Fig. 3).

To discriminate between different contexts or recognize various stimuli, each being described by many local predictors, a very large set of predictors $\{P_i\}_{i\in[\![1,N]\!]}$ might be required. Additionally, every action takes some time to perform and has a cost for the agent. In most situations,

it is thus both impossible and useless to perform all possible actions to confirm or inform each predictor. The system thus needs to select the most suitable predictors at all times, which is a difficult task when facing ambiguous or initially unknown stimuli.

2.3. Resources and activation

In order to address this issue, a quantity called *resources* is associated to each predictor. Let $R_i(t)$ be the resources for predictor P_i at time t. Let also N_i be the *necessary resources* of the predictor, i.e. the amount of resources needed for the predictor to become active, perform its action and test its prediction (run its code in more computational terms). The cost of an operation should be assigned according to its complexity and urgency, relative to the architecture and application considered[2].

We can then define the *activation variable* $A_i \in \{-1, +1\}$ as follows:

$$A_i = \begin{cases} +1 & \text{if } R_i \geq N_i \text{ (\textbf{active})} \\ -1 & \text{otherwise (\textbf{idle})} \end{cases} \tag{3}$$

As underlined earlier, prediction success is used in our approach to accumulate evidence. This is done here by updating the activation variable when predictors get activated, and this process thus does not influence idle predictors. In the end, only those that confirm their prediction remain active, by finding both their source feature in the foveated field of view before acting (time t) and their target feature afterwards ($t + dt$):

$$\overline{A_i} = \begin{cases} +1 & \text{if } A_i = 1 \text{ and } F_i^{src} \in fovea(t) \text{ and } F_i^{tgt} \in fovea(t + dt) \\ -1 & \text{otherwise} \end{cases} \tag{4}$$

If predictors cannot switch from idle to active through the previously described process, they can however obtain sufficient resources from two possible sources. The first one, called *base resources* or B_i, is defined by the user to represent the absolute relevance of the predictor, independent of time and immediate context. It might for instance bias the selection of active predictors due to external constraints (see Task 2 in the results section) or reflect habituation or learning dynamics. The other source, called *linked resources* or L_i, are the resources received from other predictors, and represents the contextual relevance of the predictor. These are predictors that previously focused on different yet congruent features that provide the necessary conditions for the considered prediction to be effectively confirmed (sequence of predictions).

2.4. Attractor Predictors Network (APN)

The Attractor Predictors Network (APN) constitutes the bulk of the architecture. It consists of a Hopfield network (Hopfield, 1982) that connects predictors and distributes resources between them, so that their activity can be well coordinated. The presence of a limited amount of resources guarantees only a small subset of predictors will be active at the same time, by putting them into competition, thus only allowing mutually congruent predictors to remain active in the end. Additionally, learning leads the system to select the most suitable predictors depending on the context and problem to solve; see also (Kokinov, 1994; Pezzulo and Calvi, 2007) for related models managing the distribution of resources among components.

[2]The possibility to assign to predictors different computational resources is not used in the tasks described below (here all predictors are assumed to have the same cost for simplicity), but this is something that we left open to show the generality of the proposed model and its possible further developments. The exact values of these costs could be learned through repeated interactions during task execution.

Table 1: Hopfield to APN terminology

Hopfield network (as in (Amit, 1989))		APN	
Neurons		Predictors (P_i)	
Symbol	Description	Symbol	Description
U_i	Post-synaptic potential	R_i	Predictor computational *Resources*
T_i	Activation threshold	N_i	*Necessary Resources* to become active
S_i	Spike	A_i	*Activation* needed to run operations
J_{ij}	Synaptic matrix	J_{ij}	Resources network's *Links* matrix

The following paragraphs describe the functioning of such a network and show how it can be altered and extended to integrate the active and normative capabilities of predictors (used as components).

We start from the mathematical definition of a Hopfield network taken from (Amit, 1989), where terms are translated in the APN terminology. Table 1 makes it possible to easily transfer known results in Hopfield networks to resources networks, and thus to APNs. Eq. 5 is another formulation of Eq. 3, only more consistent with classical Hopfield network descriptions. In turn, Eq. 6 transposes into resources the computation of the postsynaptic potential within neuron i from the excitatory and inhibitory activities of connected neurons. Synaptic weights between neurons are encoded in matrix J, with positive values corresponding to excitatory links. The original equation exactly corresponds to the computation of the linked resources, and is only altered to include the base resources, thus enabling the possibility to boost the resources associated to predictor P_i when $B_i > 0$.

$$A_i(t) = sign(R_i(t) - N_i) \tag{5}$$

$$R_i(t+1) = B_i + L_i(t+1) = B_i + \frac{1}{2} \sum_{j=0, j\neq i}^{N} J_{ij}(\overline{A_j}(t) + 1) \tag{6}$$

Another slight difference with the classical Hopfield networks' update equation lies in the use of $\overline{A_i}$ instead of A_i, in order to account for the active nature of predictors. Instead of being passively controlled by the network, they indirectly participate in its dynamics through the feedback they provide. The entire APN, considered as a resources network enriched with predictors, therefore integrates both the interactions between predictors and the interactions between predictors and the environment of the agent.

2.5. Predictors' Attractors

An Hopfield network is used to memorize patterns which are represented by the attractors of the dynamical system that it defines (Rojas, 1996). Whatever the configuration of an Hopfield network, it will sooner of later converge to one of its attractors. The set of initial configurations leading to the same attractor belongs to its *basin of attraction*.

The same attractors are found at the core of an APN and are called *Predictors' Attractors*. Each attractor $A^\mu = (A_1^\mu, A_2^\mu, \ldots, A_N^\mu) \in \{-1, 1\}^N$ is a configuration of active predictors (that will execute their operational code and should confirm their predictions) and non-active predictors (that will remain silent). Each A_i here represents the asymptotic activity of predictor P_i, i.e. its activity after an indefinitely long number of resources network updates (see Eq. 6).

Another interesting feature of an APN concerns its ability to perform well in the presence of noise or when the number of agents is high. Robustness to noise means that the retrieval of a predictors' attractor can be done even if the network is under the influence of noise. Moreover, noise can be considered as the temperature of the system and determines the trade off between exploration and exploitation. It is thus required to escape from local minima and to destabilize spurious attractors, which are a mixture of memorized attractors.

In this paper, artificial Gaussian noise is added to the predictors both to prove robustness and to enable the use of the Montecarlo method to obtain faster convergence. As a result, the probability of having the resources associated with predictor P_i equal to R is:

$$Pr(R_i = R) = \frac{1}{\sqrt{2\pi\delta^2}} exp[-\frac{(R - \bar{R}_i)^2}{2\delta^2}] \tag{7}$$

where $\bar{R}_i$ is the statistical mean of the random variable R_i. The distribution variance δ and the temperature T are related as $T = (2\sqrt{2}\delta)^{-1}$ (for a similar definition of noise in Hopfield networks see (Amit, 1989), pages 66-67).

In the end, predictors' attractors represent long-term multimodal frames of predictors used to achieve representation, categorization and motor control. They not only are a collection of several features, but they also have active capabilities. They are characterized by a coordinated pattern of activity and strong associative links. A predictors' attractor thus arises when a set of feature-specific predictors cooperate to find regularities inside the visual input.

2.6. Learning

For decision-making and categorization tasks, we want each attractor to be associated with one category of stimuli/contexts. After choosing adequate predictors to test hypotheses through interaction, we still need to learn the link matrix that determines the dynamics within the resources network. In the more restricted context of active vision, each predictor is specialized on a specific saccade movement that shifts the fovea from a particular feature to another. So, if we assume that a category is characterized by a set of spatial relations among features, the activation of a subset of predictors (i.e. a predictors' attractor) can identify a category. The main idea of our proposal is to represent perceptual categories by a set of anticipatory active processes. Using dynamical system like representations (attractors as categories) makes it possible to study the dynamics that leads the categorization process to reach a decision, and not only its final result.

To learn the connection weights we used a procedure based on a Hebbian rule. In the tasks we will present later in this paper, we use four fixed categories to determine whether a stick figure (input image) represents a giraffe, a horse, a cat or a dog. Additionally, we started from a large database of pictures already categorized by humans in a previous experiment by Sanborn et al. (2010) (see Fig. 3 for an example), so that supervised learning can be used.

In each picture we use a brute force approach to generate all potentially correct predictors, which are all triples made of an existing starting feature, saccade and target feature that matched the corresponding stick animal. Then we defined four predictors' attractors $\{A^{CAT}, A^{DOG}, A^{GIRAFFE}, A^{HORSE}\}$, each of them composed by all predictors that were previously extracted from figures of the associated category. A Hebbian rule is generally used to memorize a set of arbitrary patterns in a Hopfield network, and works as follow: for each known attractor $A^\mu \in \{+1, -1\}^N$, with $\mu \in \{CAT, DOG, GIRAFFE, HORSE\}$ here, the resources network's weights, initialized to zero, are modified by adding the term $\Delta J_{ij}^\mu = A_i^\mu A_j^\mu$. Integrating over all attractors, the final weight's are given by:

$$J_{ij} = \begin{cases} 0 & \text{if } i = j \\ \frac{1}{N} \sum_{\mu} A_i^{\mu} A_j^{\mu} & \text{otherwise} \end{cases} \tag{8}$$

2.7. Algorithm and behavior

Once we have memorized the categories as predictors' attractor, we finally need a procedure to recall them in order to categorize any stimulus. Remembering that Hopfield networks are associative memories, we know that predictors' attractors will be reached according to their similarity with the current configuration. The algorithm used to alternate and combine the updates on the predictors and on the resources network is reproduced in Algorithm 1. It especially describes how the system dynamically converges to an attractor by selecting adequate saccades, testing the predictions and spreading resources between predictors. You can also refer to Fig. 2 to see how the different components of the model are organized and interact. Moreover, Table 2 reports the list of parameters involved in the system, with the values used in the simulations described in the following section.

To simplify the following explanations, assume that each predictor P_i has the same computational cost and same base resources. We will thus arbitrarily take $N_i = 0$ and $B_i = 0$ for all predictors. We then initialize the system by putting the fovea in a random position and by activating a small set of a predictors randomly chosen to be active for the first iteration (lines 1-2 of Algorithm 1). Active predictors have to pass three steps in order to remain active in the following iteration and thus continue on contributing to the basin of attraction of the predictors' attractor that will finally be selected:

1. Each active predictor must have perceptual relevance, which here means that its starting feature must be immediately available in the current foveal view. If this condition is satisfied, the predictor will be retained for the next step, else it will become idle (in lines 5-11 of Algorithm 1 these relevant predictors are collected in the set RP).
2. Each perceptually relevant predictor proposes its saccadic motor command to be selected by the system. The fovea then merges these potentially asynchronous commands into the next effective fixation point. A Winner-Take-All strategy is used to select the saccade with more proponents, each predictor voting for its saccade (lines 16-18 of Algorithm 1). In a case of parity the tie is broken randomly. Note that more relevant predictors can propose the same saccade (although they might share their starting feature and proposed command, the target feature can still differ). So also the selection of a single action can imply the further verification of more than one predictor.
3. Each relevant predictor which had its saccadic motor command selected finally tests if its target feature appeared, as a result of the fovea movement. If the matching occurs, the prediction was fully successful and the predictor is kept active to contribute to the definition of a basin of attraction during the following iterations (in lines 19-24 of Algorithm 1 these successful predictors are collected in the set SP). Furthermore if this happens, other predictors of the same category are activated ($A_i = +1$ for other s predictors, where s is called *boosting parameter*), enriching the basin of attraction in order to boost the evidence collected for faster convergence (lines 25-28).

At the end of each iteration, the fovea moved to a new position and predictors which have seen their commands selected and predictions satisfied are active. Moreover, related predictors are also activated according to the spread of resources within the network and a few more are randomly flipped on or off by the temperature simulated in the system (lines 40-41 of Algorithm 1). The success of a predictor pertaining to a given category therefore has a positive impact on predictors that respond to the same category (remember that predictors from the same attractor are linked

Algorithm 1 Categorization by the APN

// *Randomly position the fovea and activate a predictors*
$fovea_{pos} \leftarrow$ random position
$\{A_i\} \leftarrow$ random vector in $\{-1,+1\}^N$ where $card(A_i = 1) = a$
repeat
  $RP \leftarrow \emptyset$ // *Relevant predictors*
  for all P_i **do**
    // *Only consider active predictors with adequate context*
    if $A_i = +1$ and $F_i^{src} \in fovea_{view}$ **then**
      $RP \leftarrow RP \cup P_i$
    end if
  end for
  // *Try saccades until at least one prediction is confirmed*
  $\{\overline{A_i}\} \leftarrow \{A_i\}$
  $SP \leftarrow \emptyset$ // *Successful predictors*
  repeat
    // *Perform the saccade with the maximum number of votes*
    $S_{best} \leftarrow argmax_S\{card(S_k = S | P_k \in RP))\}$
    $fovea_{pos} \leftarrow fovea_{pos} + S_{best}$
    // *Test all relevant predictors*
    for all $P_j \in RP$ **do**
      // *Confirm those where action and prediction both match*
      if $S_j = S_{best}$ **then**
        if $F_j^{tgt} \in fovea_{view}$ **then**
          $SP \leftarrow SP \cup P_j$
          // *Activate s predictors of the same category*
          for s randomly chosen $P_v \mid \forall \mu\ A_v^\mu = A_j^\mu$ **do**
            $A_v = 1$
          end for
        else
          $\overline{A_j} \leftarrow -1$
        end if
        $RP \leftarrow RP \backslash P_j$ // *No more selection of this predictor*
      end if
    end for
    // *Revert the saccade if no predictor was successful*
    if $SP = \emptyset$ **then**
      $fovea_{pos} \leftarrow fovea_{pos} - S_{wta}$
    end if
  until $SP \neq \emptyset$ or $RP = \emptyset$
  Update the network with Montecarlo method at temperature T
  $\{A_i\} \leftarrow$ application of Eq. 5, 6 & 7
until $\exists \mu \mid \frac{card\{A_n=1 \text{ and } A_n^\mu=1\}}{card\{A_m^\mu=1\}} > t$ // *A category has frequency of active predictors* $> t$
repeat
  $\{A_i\} \leftarrow$ application of Eq. 5, 6 & 7
until $A \in \{A^\mu\}$ // *Predictors' attractor reached*
return μ // *Return the associated category*

with strong positive weights). This improves the efficiency and robustness of the categorization process. These steps must however be repeated, moving the fovea in each iteration, until enough evidence is collected. More specifically, a scan path should collect evidence about which predictors deserve to be part of the basin of attraction, because this is what determines the convergence to a predictors' attractor instead of another. Tested successful predictors should be kept active across the iterations and unsuccessful ones should become idle. The role of $\overline{A_i}$ is to keep track of relevant unsuccessful predictors (i.e., those that matched with the starting feature but not with the target one). Considering that these predictors are more likely to represent a wrong category through this variable we inhibit their ability to contribute to the activation of other predictors of the same category during the network update. As it is possible to see from Algorithm 1, $\overline{A_i}$ starts as copy of A_i (line 13) and then, if some relevant unsuccessful predictors are encountered, to some of its elements can be assigned an idle value (line 30). Note that other evaluated predictors have already $\overline{A_i}$ equal to +1, because only active predictors are taken in consideration. This is the reason why in Eq. (6) the current resources are computed over the $\overline{A_i}$ variables: to integrate the action outcome within the network dynamics.

To automatically understand when enough evidence is collected we chose to define a threshold parameter (called t) that indicates when the basin of attraction is rich enough to properly recall a predictors' attractor: when one of the categories reaches a number of active predictors, which divided by the cardinality of its predictors is greater than t , the architecture stops to saccade (line 42). Starting from (Amit 1989), who already investigated the relation between recalled attractors and basins of attraction, we found that $t = 0.3$ worked fine for the tasks described below. Then, once all the saccades are performed, the resources network uses the knowledge stored in its links to enrich the evidence collected by predictors through the last updates until convergence (lines 43-45). Finally, when a predictors' attractor is reached the decision is made (line 46).

3. Simulation results

To validate the proposed architecture we tested it in three categorization studies. In the first study, the architecture was trained to recognize figures of four categories (CAT, DOG, GIRAFFE, HORSE) and tested in a categorization task with exemplar and morphed stimuli. As we can see in this example little feature variations can lead to different categorical decisions. In the second study, we modified the probability of executing different eye movements so as to study how sensorimotor and active vision processes modify the choice. Finally, we studied the relationship between the morphing parameters and the accuracy of the system, with a qualitative comparison to biological data.

3.1. Task 1: Categorizing through attractor dynamics

We trained our system on a learning set of 32 pictures originally developed by Olman and Kersten (2004) and categorized in Sanborn et al. (2010). As previously said, in each picture we considered all the possible triples (starting feature, saccade, target feature) in order to build all the possible predictors. At the end of this process we obtained 1176 predictors [3]. Among these predictors, 224 were extracted from pictures of the category CAT, 319 from pictures of the category DOG, 319 from pictures of the category GIRAFFE and 314 from pictures of the category

[3] A number that depends on the chosen offsets used to distinguish feature descriptors and saccades belonging to different predictors.

Table 2: APN parameters and their values used in the simulations.

Parameter	Description	Value
N	Total number of predictors	1176
B_i	Predictor Base Resources ($\neq 0$ only for simulation of Figure 7)	0.23
M	Number of Gabor filter's orientations	16
T	Temperature (noise) of the system	0.2
a	Number of initial active predictors	30
s	Boosting parameter	75
t	Threshold frequency of category active predictors reached to stop scanning	0.3

HORSE [4]. After the four predictors' attractors have been obtained through the application of the Hebb Rule (Eq. 8), we used Algorithm 1 to categorize a test set of new pictures with $Temperature = 0.2$ and 30 randomly chosen predictors activated in the first iteration ($a = 30$). As reported in the algorithm, the Montecarlo method was used to speed up simulations.

The architecture was validated with a test set of 32 stick animal figures, each one processed five times to amortize the fluctuations produced by the Temperature's noise and the many possible scan paths. The categorization was correct 120 times, so we had 75% of correct answers. Furthermore, the system used an average number of saccades equal to three to perform these categorizations, which is comparable to humans. The performance of other systems was: Linear Discriminant Analysis: 87%, Gaussian NaiveBayes Classifier: 90%, Decision Tree: 68%. These classifiers were trained on vectors composed by nine float values that uniquely identify a stick animal (see (Olman and Kersten, 2004)). These values represent the feet angles, body height, body angle, tail length, tail angle, neck length, neck angle, head length and head angle of the corresponding animal. Note that these classifiers were directly trained using these vectors where as our system processes only the visual input through the described feature descriptors. The same learning and test phase structures were used with both these classifiers and our system.

We report below two plots that show the dynamics of the categorization process (Fig. 4 and Fig. 5). The following overlap measure between the resources network's configuration at time t and a predictors' attractor is defined:

$$m^{\mu}(t) = \frac{1}{N}\sum_{i=1}^{N} A_i^{\mu} A_i(t) \qquad \mu \in \{CAT, DOG, GIRAFFE, HORSE\} \tag{9}$$

The overlap is a real number with $m^{\mu}(t) \in [-1.0, 1.0]$. Its maximum value is 1.0, which corresponds to an identity between the current configuration and the predictors' attractor μ. On the other hand the overlap's minimum value is -1.0, meaning that the the configuration at time t and the predictors' attractor μ are opposite.

In Fig. 4 and Fig. 5 we report the evolution in time of the overlap between the resources network's configuration and the four predictors' attractors CAT, DOG, GIRAFFE and HORSE during the categorization process. In certain time steps we also showed the starting and target features of the successful predictor which took control of the fovea at that time. This was done in order to show the dependence of the categorization's dynamics on the current foveated features.

[4]There are fewer predictors belonging to the category CAT because the corresponding stick animals are smaller and thus produced less saccades' variability among the different predictors.

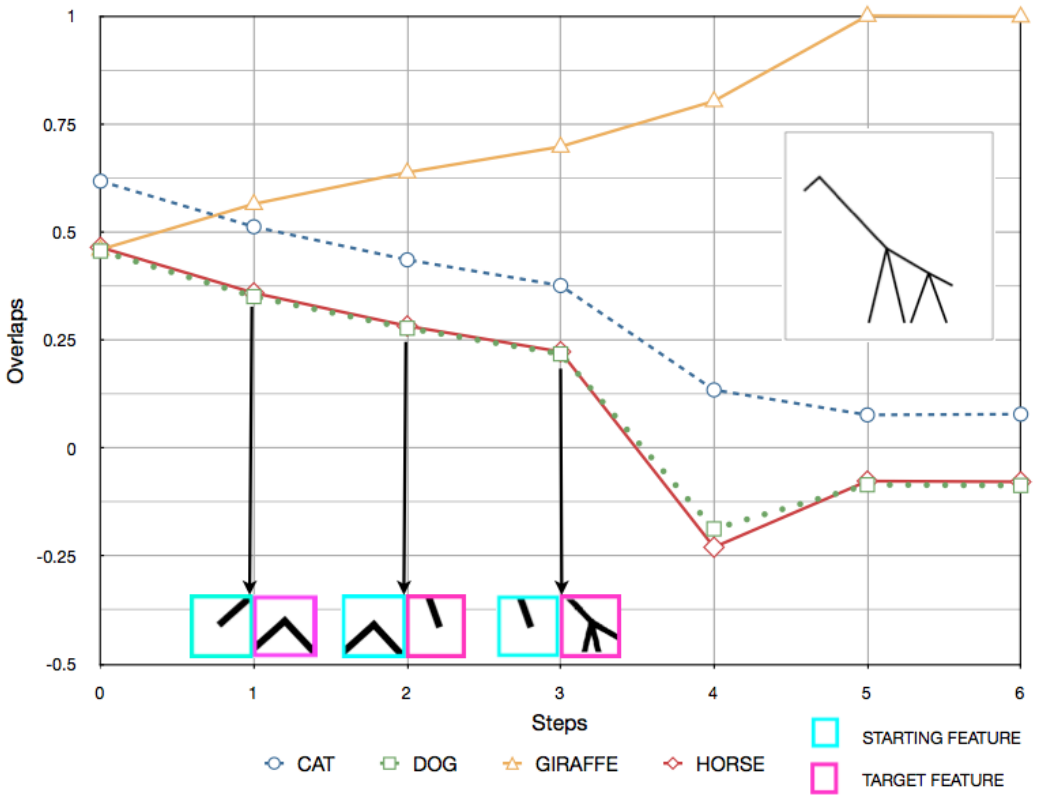


Figure 4: Overlaps of the predictors' attractors CAT (green), DOG (blue), GIRAFFE (magenta) and HORSE (red) with the resources network's configuration at each time step. Every time step we reported the starting feature (cyan) and target feature (fuchsia) of the successful predictor which took control of the fovea. In this example 3 saccades were necessary to build a rich enough configuration that led the resources network to recall the proper predictors' attractor GIRAFFE, ending with an overlap close to 1.0. In this example we have a stimulus that can be considered unambiguous, because the predictors' attractor GIRAFFE dominates the others since early stages of the categorization process.

Indeed, as we will see, two different scan paths made on the same stick animal can bring two different categorization's dynamics and sometimes also to the recognition of different categories.

Fig. 4 shows a successful categorization where one predictors' attractor dominates the others from the beginning. On the other hand, Fig. 5 shows the categorization of a more ambiguous stimulus that makes predictors' attractors compete before a final decision can be taken. Overall, the results of this simulation show that the proposed model achieves good classification performance. Compared to other classification systems, the proposed method aims at reproducing key dynamic and competitive aspects of the choice (see Fig. 4 and Fig. 5). The comparison between the two situations shows that the model is sensitive to the amount of perceptual ambiguity. In a situation with low uncertainty, the competition is resolved very early, whereas with high ambiguity it requires more time, coherent with psychological models of drift-diffusion (Ratcliff, 1978; Ratcliff and Rouder, 1998).

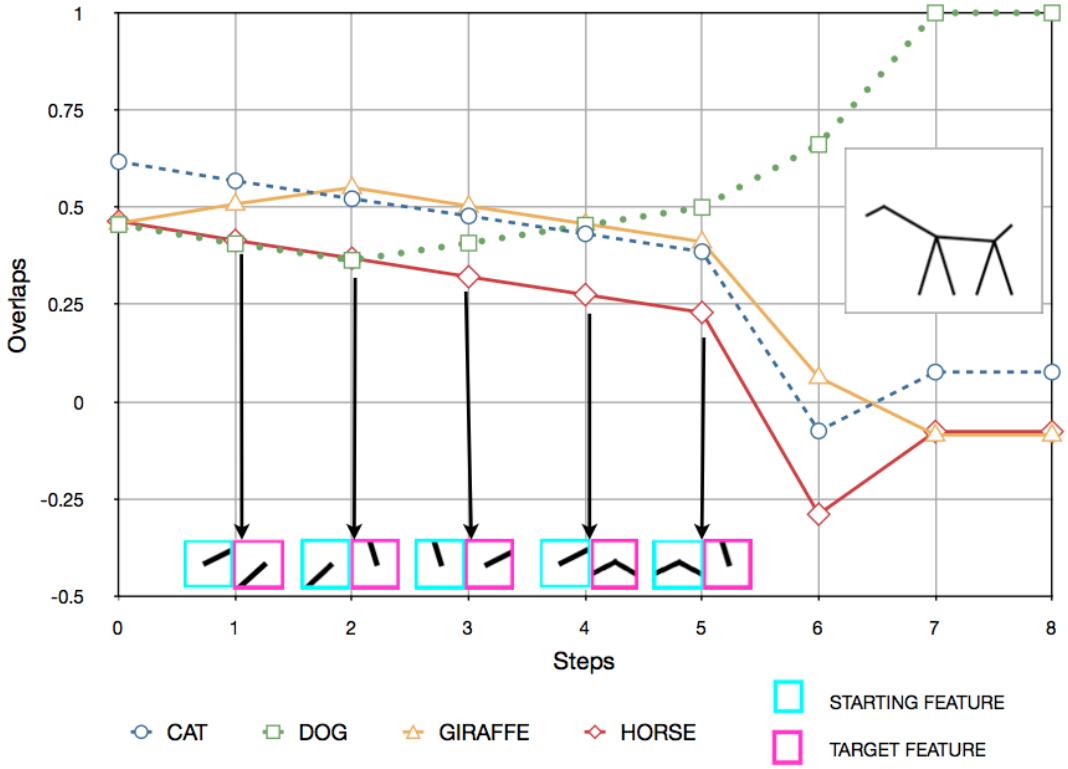


Figure 5: Categorization dynamics made on a stick animal categorized by humans as a DOG. Note how the ambiguity of this stick figure keeps the overlaps close to each other until a final decision is taken, finally making the system categorizes it as a DOG. Here more ambiguity implies also a larger number of saccades. Note also how in the last two steps the network starts to repeat its configurations (i.e., reaches a predictors' attractor) when the choice is made.

3.2. Task 2: How sensorimotor processes and active vision influence choice

To observe how different scan paths made on the same stick animal can bring diverse categorization results, we can compare the two categorization's dynamics represented in Fig. 5 and Fig. 6 where two different saccadic sequences are executed on the same stick animal (a horse), but yet lead to distinct categorization outcomes (DOG and HORSE). A study conducted by Kietzmann et al. (2011) reported a similar effect: the experimenters were able to bias the categorization of ambiguous figures by influencing the subjects' scan-paths, and concluded that the actions (eye movements) are constitutive of perception.

To better understand how this result emerges in our model, we now discuss in detail a categorization example. We chose the ambiguous stick animal shown in Fig. 7. As we can see, it is not easy to determine a priori what stick animal it represents, indeed our architecture sometime categorizes it as a CAT and sometimes as a DOG. To underline how much different scan paths can bring the categorization process to different results we will manually augment the probability of activating some predictors instead of others. To implement this idea we can increase the base resources B_i of these predictors (that before was assumed to have a value zero giving the same absolute relevance to all predictors). Considering the presence of the noise produced by the system's Temperature, this will increase the probability of doing saccades among some chosen

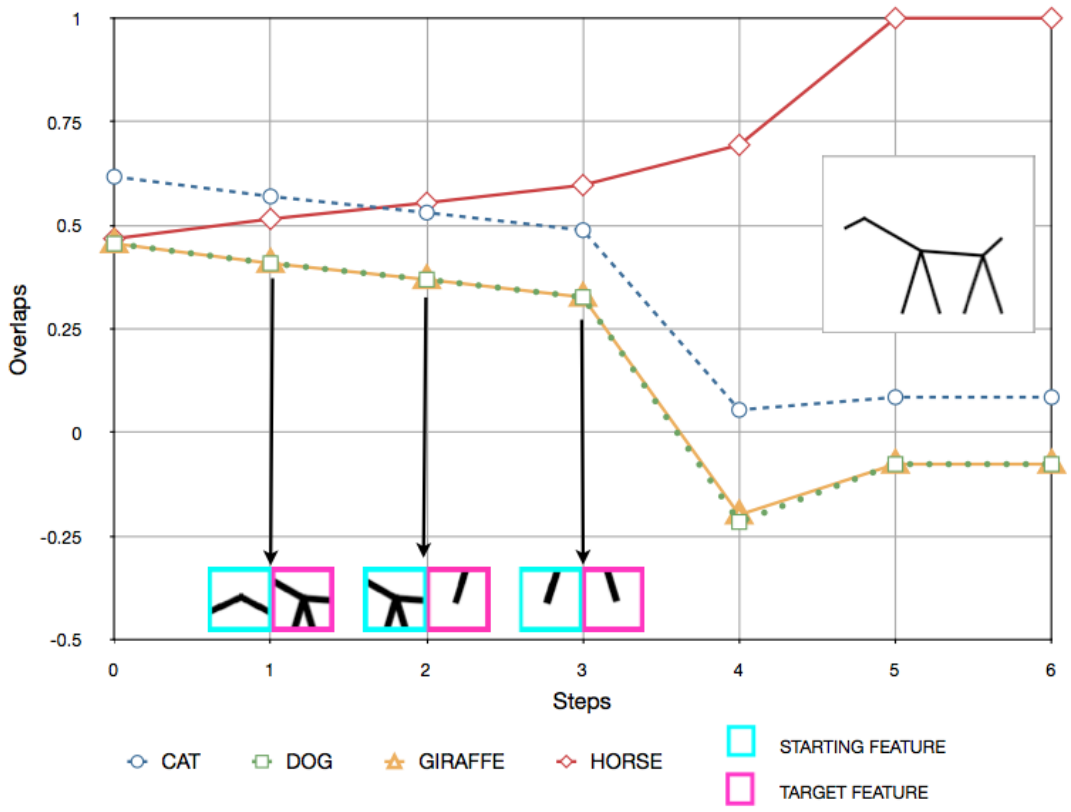


Figure 6: In this figure we reported the categorization dynamics made on the same stick animal as in Fig.5 but with a different sequence of foveated features. As we can see this brought the architecture to recognize a HORSE instead of a DOG. So the category of a stick animal is not absolute and strictly depends on the sequence of the foveated features.

features.

As we can guess from Fig. 7, the set of boosted predictors has a strong influence on the categorization results. Indeed if we increase the base resources B_i of predictors belonging to the category DOG the ambiguity is broken and the stick animal is categorized as a DOG; symmetrically if we increase the base resources of predictors belonging to the category CAT the stick animal is categorized as a CAT.

Among ten different scan paths if $B_i = 0 \quad \forall i \in [1, N]$ we found a mixture of DOG's categorizations and CAT's categorizations. This is reflected in Fig. 7 by the average overlaps represented in blue (DOG) and green (CAT) with circle data points. On the contrary if the number of predictors belonging to the category DOG, with an assigned value of base resources B_i equal to b, is greater than c ($| \ i \in [1, N] \ s.t. \ A_i^{DOG} = 1 \ and \ B_i = b \ |> c$), where we set $c = 7$ and $b = 0.23$ [5], the system categorizes DOG in all the ten runs, as we can see from the

[5]Here the value 7 for c and the value 0.23 for b were chosen manually to show that a small number of predictors with a little additional amount of activation are enough to change the overall system behavior.

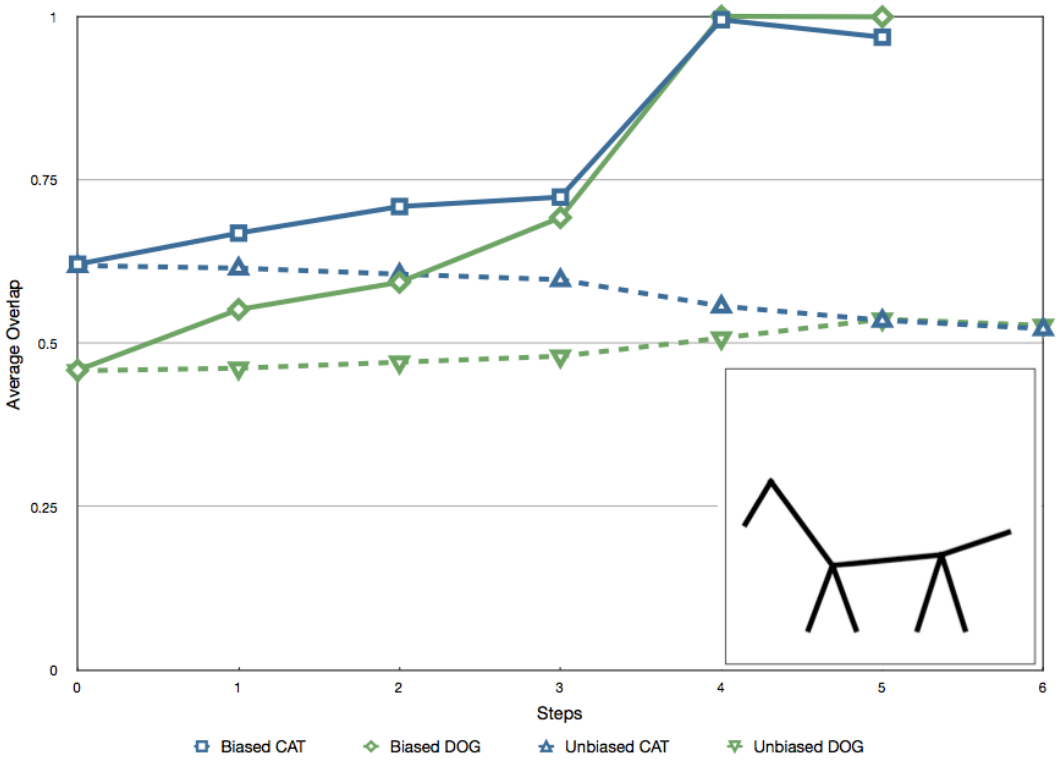


Figure 7: Average overlaps over ten categorizations of the stick animal reported above. Average dynamics of the unbiased system with $B_i = 0$ for all predictors represented together with the dynamics with the biased system with $B_i \neq 0$ for some predictors. As we can see augmenting the probability to visit some features instead of others has a strong influence on the categorization process. Note that the unbiased categorizations for this stick animal requires to perform an additional saccade.

average overlaps represented with square data points in Fig.7 (the case of the category CAT is the same as the DOG case). Although this is not reported in detail here, the same test was performed on many other ambiguous stimuli and yielded the same results. Overall, the results of this simulation show that the active perception loop can influence the categorization process; by eliciting different action-outcome sequences, evidence is accumulated in different ways and the system can fall in distinct attractors. This implies that different sequences of saccades might lead to different categorizations, as shown experimentally (Kietzmann et al., 2011).

3.3. Task 3: How morphed stimuli influence choice

Akrami et al. (2009) report a study involving two macaques performing a 2-alternative-forced-choice delayed-match-to-sample (2AFC-DMS) task, consisting in categorizing a set of pictures. In each trial they started from two different pictures, among a complete set of stimuli, and produced a set of morphed variants of them. The monkeys' task was to choose the most similar picture to the morphed variants among the two original pictures. A linear relation between the number of times the monkeys chose one of the original pictures, taken as reference category, and the morphing parameter used to produce the stimuli was found (except for the extreme morphing values leading to variants close to the original images).

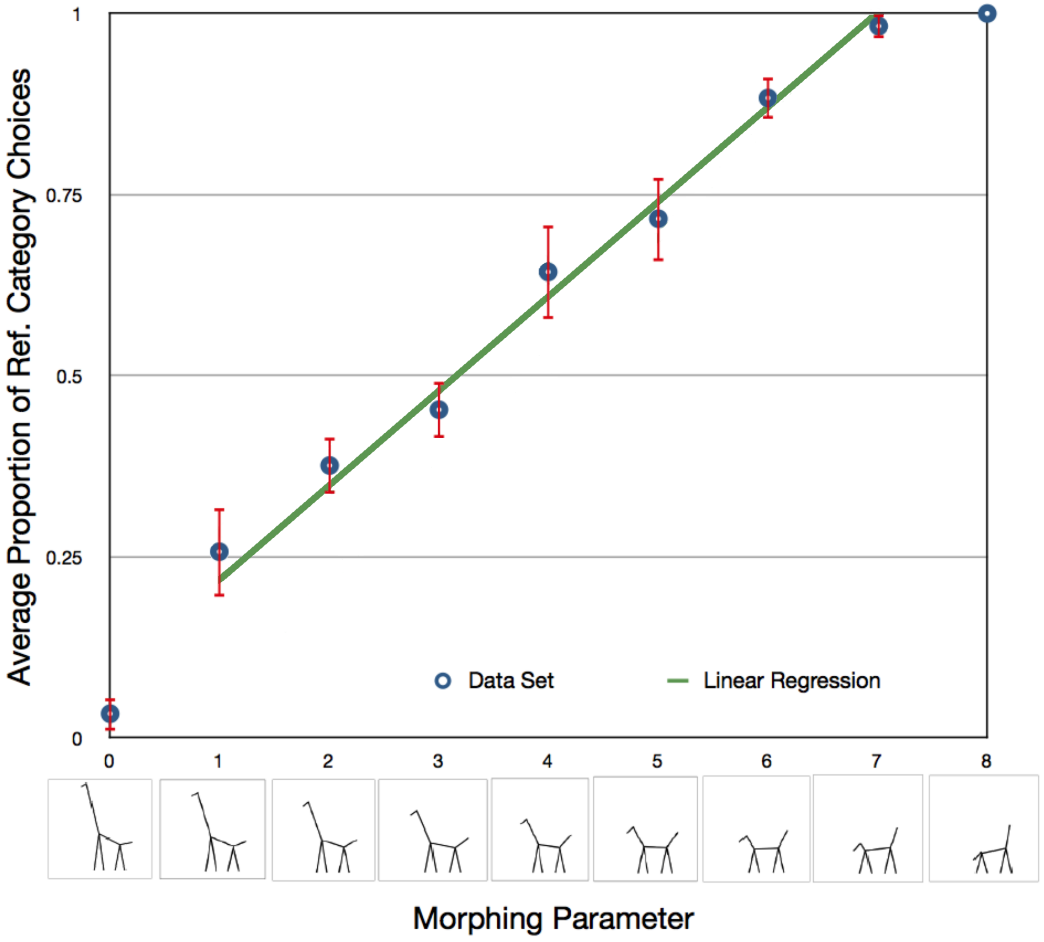


Figure 8: Result of the linear regression computed for the proportion of times the system chose the reference category (morphing parameter 8) as a function of the morphing parameter. The slope of the line is 0.149. Error bars represent mean values and standard deviations for the 60 simulations performed. The series of morphed stick animals is arbitrary and simply illustrates the morphing.

To underline the cognitive plausibility of our proposal we tested our architecture on a very similar task. We randomly chose 12 pairs of our stick animals and produced 7 morphed variants for each of them. The original stick animal taken as reference has a morphing parameter 8 and the other original stick animal has a morphing parameter 0. All the remaining values between these two extremes represent the corresponding morphed variants of the two original stick animals (see Fig. 8). For each of the 12 pairs we performed 5 simulations where our architecture categorized the 9 possible images (the 2 originals plus the 7 variants). In each trial, instead of keeping the four possible predictors' attractors memorized as before, we memorized only the two predictors' attractors representing the two target categories. Using linear regression we found the same kind of linear relation as in (Akrami et al., 2009) between the average proportion of times our architecture chose the reference category (among 60 simulations) and the values of the morphing parameter.

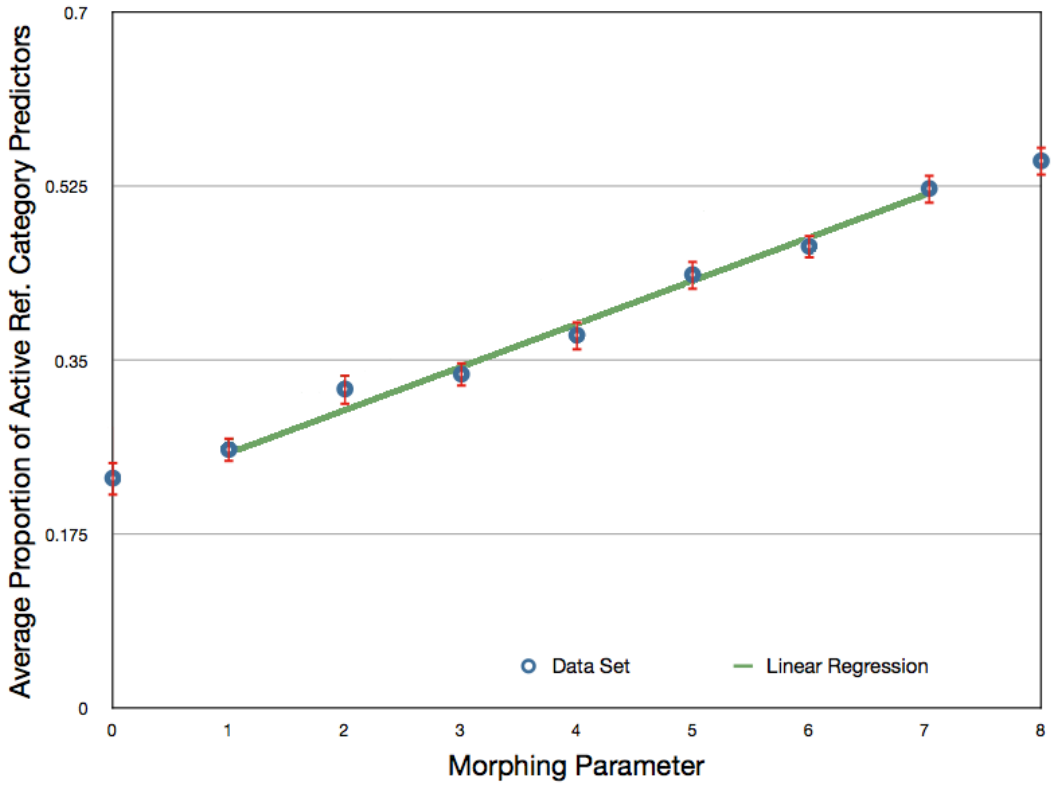


Figure 9: Result of the linear regression computed for the proportion of active predictors belonging to the reference category (morphing parameter 8) as a function of the morphing parameter. The slope of the line is 0.043. Errors bars represent mean values and standards deviations for the 60 simulations performed.

Akrami and her colleagues also recorded the neural activity of the macaques' neurons with an effective response to the reference category. A linear relation between the neural population responses, averaged across trials, and the morphing parameter used to produce the stimuli was found also in this case (still taking apart the extreme values close to the original images). We found a similar result measuring the proportion of active predictors belonging to the reference category averaged among 60 simulations. The resources network configuration considered here is the one before a predictors' attractor is reached. As we can see from Fig. 9 there is a linear relation between the average proportion of active predictors belonging to the reference category and the morphing parameter. Overall, the results of this simulation show that, similar to the study of Akrami et al. (2009), there is a linear relation between the morphing parameter and the response of the predictors in the proposed model.

Although our task is similar to the one described above, there are some differences between the behavioral task performed by the monkeys and the task performed by our architecture. First our task is not a 2AFC-DMS, in fact only one stimulus to categorize is used as input in each trial. This implies that in our case there is no decision about the similarity of two images because in each trial the category of a single image has to be determined. Then the role of the two original images used in one trial of the task performed by the macaques is asymmetrical. Indeed, they recorded the activity of neurons that were giving an effective response to only one of the two

images (defined as "Eff image" in Akrami et al., 2009). In our task instead the role of the two original images used in one trial is fully symmetric because both corresponding predictors' attractors are stored before the architecture performed the task.

In their work an attractor neural network is presented to model the behavioral task performed by the two monkeys. Although our resources network is a Hopfield-like network too the attractor predictors network adds other interesting features compared with classical neural networks. One is the role played by prediction, implemented in the predictors' activities. Moreover a visual attention controller is distributed among these competing predictors. Finally their neural network stores uncorrelated random patterns. On the contrary in the resources network the stored predictors' attractors are learned from a set of stimuli and they represent the long term activity of successful predictors for a certain category.

3.4. Further analyses: how categorization depends probabilistically on the choice of saccade sequences

In the proposed system, the likelihood of a categorization depends on the sequence of saccades selected to visually explore the object. In more formal terms, one can assign to each saccadic sequence a probability that it will lead to the recognition of a certain category (see Fig. 10). Indeed saccadic sequences are not all equivalent because, as we observed in our simulations, the features' triples (F^{src}, S, F^{tgt}) of predictors can be less or more informative. Triples that are represented by predictors belonging to all the categories can be considered as non informative since their application can support all the hypotheses at the same time, thus not contributing to one category in contrast to the others. On the other side the application of a triple that is codified by predictors belonging to only one category is clearly supporting the recognition of only that category. To give an example taken from our experiments we found triples codifying a very long neck (head and torso features linked by a very long saccade) represented by predictors belonging only to the GIRAFFE category. On the other hand we found the legs or the tail to be less informative.

To fully characterize the observations made above, on the basis of all the simulations described here, we computed such probability distributions to exploit the relation between predictors involved during a scan path and the category recognized when it is executed. As we can see in Fig. 10, the probability that a saccadic sequence will bring the system to choose a certain category is proportional to the number of predictors that belong to that category involved in the categorization process. Fig. 10 shows that when more that one third of the predictors that control a scan path belongs to a certain category, that category has a high probability to be chosen by the system. The results of this analysis are compatible with active perception and sensorimotor theories that link categorization to "the mastery of the sensorimotor contingencies" that govern object exploration (O'Regan and Noe, 2001). In the proposed system, sensorimotor contingencies can be conceptualized in probabilistic terms as sequences of features triples (F^{src}, S, F^{tgt}); see also (Maye and Engel, 2011). Our results show that (so-defined) sensorimotor contingencies play an important role for the categorization. Further analyses are necessary to assess how much information needs to be conveyed by sensorimotor contingencies for determining a categorization, and if and when there is a phase structure in the categorization (e.g., in Fig.10 0.3 seems to be a critical value).

4. Conclusions

We presented a computational theory of perceptual categorization that elaborates on the widespread idea of evidence accumulation but includes elements of grounded and sensorimotor

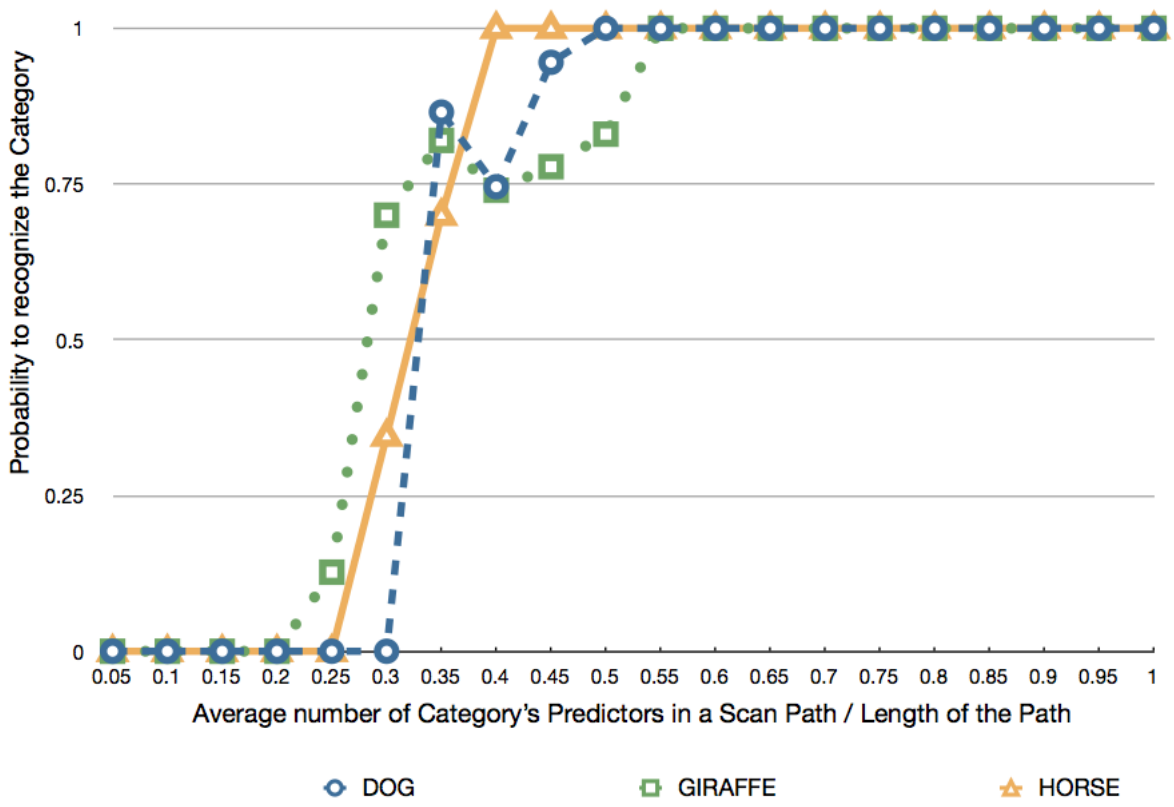


Figure 10: In this figure we reported the probability to recognize a certain category as a function of the frequency in a scan path of that category's predictors. To compute this probability we processed data from all the 160 simulations performed to evaluate the success rate of the architecture. To obtain the frequencies we divided the number of predictors of a certain category that produced a scan path by the length of the path. Then to compute the corresponding probability we analyzed how many times a certain frequency brought to the recognition of a category instead of another. We observe that the probability that a saccadic sequence will bring the system to choose a certain category is proportional to the number of predictors that belong to that category involved in the categorization process.

theories of cognition to determine what is the conceptual content (and counts as evidence) for the choice, and how this information is elicited during the decision.

The proposed model is based on three key hypotheses: (1) action dynamics shapes the ongoing categorization; (2) evidence accumulation and action guidance are both based on prediction success; (3) evidence accumulation follows attractor dynamics that mimics structured agent-environment interactions. The specific implementation of the proposed model is based on a network of sensorimotor predictors that encode category-specific regularities and permit re-enacting them during visual exploration. Hopfield networks were selected as models of this process given their prominence in modeling memory dynamics (Amit et al., 1997; Miyashita and Chang, 1988; Sakai and Miyashita, 1991).

The model retains essential aspects of diffusion-to-bound (Ratcliff, 1978; Ratcliff and Rouder, 1998) and related models (Lepora et al., 2013; Usher and McClelland, 2001); still it has several elements of novelty. In particular, the proposed model incorporates sensorimotor dynamics that

are disregarded in other approaches. While in diffusion-to-bound models what is accumulated over time is a sensory evidence, in the proposed model what is accumulated are the category-specific predictions, which in turn are generated by the re-enactment of action-outcome pairs associated to the category. Furthermore, while in diffusion-to-bound models evidence is accumulated passively, the proposed model uses active perception to select the next sensory evidence.

The simulations we performed show that the model can achieve good performance while at the same time reproducing key aspects of perceptual decisions such as an appropriate sensitivity to ambiguity and other characteristics of the stimuli (e.g., morphing coefficients), and a dependency of the categorization from the sensorimotor contingencies elicited while exploring the to-be-categorized stimulus. Our model assumes that when possible active vision is used to select the next saccade and the next stimuli to attend, as the active selection of stimuli is plausibly an important aspect of naturalistic choices (Ballard, 1991). Even so, we recognize that the active selection of stimuli is not a necessary requirement and it is certainly possible to recognize categories without doing any saccades (for example, when objects are presented briefly). Recognizing a category without doing saccades is possible in our model by activating several (or even all) predictors that link to the currently available portion of the image and without using the active vision part of the model, which would plausibly imply a decrease in the model performance.

An open objective for future research is refining the active perception process so that it elicits the most relevant and informative evidence. Indeed, in the proposed model not all interactions are informative for a decision, but only those that lead to bifurcations of the categorization dynamics, and future extensions of the model should consider this aspect. Another objective for future research is applying the method to more complex choice scenarios by incorporating realistic feature descriptors and by studying how sensorimotor patterns can encode the hidden structure of complex objects. Furthermore, the model can be extended by using interoceptive and affective codes (rather than only perceptual and motor codes) that are elicited during the interaction with objects in the same way perceptual and motor codes are (Barsalou and Wiemer-Hastings, 2005; Pezzulo, 2013).

A final objective for future research is exploring the putative brain correlates of the proposed model. This is not an easy task, since the attractor predictors network is a hybrid architecture that combines different mechanisms. In fact the resources distribution system uses a neural model (an Hopfield network), while predictors have a more procedural nature that encodes simple sensorimotor computational modules, which are not directly implemented as a neural network. Despite that, a characteristic of the proposed model is that active perception and decision processes are tightly linked. Recent research on perceptual choices reveals that the neuronal circuits involved in saccade and attention control (e.g., the lateral intraparietal area of monkeys) are also involved in decision-making and in particular they show neuronal signatures of evidence accumulation for the choice alternatives (Gold and Shadlen, 2001, 2007), even in tasks requiring the acquisition of novel arbitrary associations between visual stimuli and actions (Tosoni et al., 2008). Further evidence supports a continuous flow model according to which evidence is continuously feed to and used by motor areas even before the decision is completed (Coles et al., 1985). This body of evidence suggests that perception, decision and action are not staged and segregated processes, which is consistent with our assumption that categorization requires continuous interactions between them. It remains to be tested if the bidirectional interactions between categorization dynamics and the choice of action movements (e.g., saccades) postulated in our model are realistic and which brain networks support them.

Another characteristic of the proposed model is that it uses predictive information elicited by action-outcome pairs for the categorization. A series of studies have demonstrated that the visual appearance of objects primes motor programs and activates neural responses in motor areas of the brain (Chao and Martin, 2000; Tucker and Ellis, 2004); these activations could provide

the necessary predictive information for our model to function. Our model would predict that categorization should be impaired if these automatic motor responses are suppressed. Evidence exists that interfering with the perceiver's motor activity influences the recognition of actions performed by others (Kilner et al., 2003) and can influence the categorization of ambiguous figures (Kietzmann et al., 2011) but further studies are necessary to fully assess the role of action prediction mechanisms in categorization.

4.1. Linking grounded and sensorimotor theories of cognition with dynamic models of choice

Another novelty of our model is that it reconnects grounded and sensorimotor theories of cognition with dynamic models of choice. A prominent view in grounded cognition is the *Perceptual Symbol Systems (PSS)* theory, which is based on the two concepts of *perceptual symbols* (*records of the neural activation that arises during perception* (Barsalou, 1999, p. 583)) and *simulators* (multimodal frames of perceptual symbols which afford motor control, representation, and categorization). In this framework, categorization consists in the (partial) re-enactment of simulators and its associated perceptual symbols, which in turn re-creates the experience associated with the original acquisition of the concept while at the same time situating it in the current sensorimotor context (this is why it can be referred to as a *situated simulation* (Barsalou, 2003)). This theory provides a coherent framework that also somewhat subsumes prototype and exemplar theories, because both can emerge from the dynamics of situated simulations. Furthermore, it also explains ad-hoc categories (e.g., fruits vs. things that could be useful for camping) that can be formed de novo based on an agent's goals (Barsalou, 1983).

The original *PSS* theory (Barsalou, 1999) focuses principally on the importance of multimodal perceptual codes for categorization; more recent developments of the theory also highlight the importance of action codes and predictions (Barsalou, 2009; Pezzulo et al., 2011) as well as of interoceptive and affective codes, especially for processing abstract concepts (Barsalou and Wiemer-Hastings, 2005; Pezzulo, 2013). This implies that according to PSS the architecture of a concept is constituted by a multifarious set of modal codes. However, the theory does not fully specify which elements are stored in the perceptual symbols and what are the mechanics of re-enactment and situated simulations (e.g., during a perceptual categorization task). Our model proposes that the stored perceptual symbols largely correspond to action-outcome links that are (typically) elicited during situated interactions, and simulators organize them into (longer) sequences that encode characteristic patterns of multi-modal interactions, and which in our model are linked to predictors' attractors.

We also argue that the dynamics of simulators are linked to two key phenomena. On the one hand, simulators follow attractor dynamics, where attractors are essentially a powerful mechanism for encoding and re-enacting category-specific information (Pezzulo and Calvi, 2011). On the other hand, situated simulations are guided by active perception dynamics that elicit task-relevant information rather than collecting it in a purely bottom-up way (Ballard, 1991; Hayhoe and Ballard, 2005; Rothkopf et al., 2007; Yarbus, 1967). As a result, our model fuses elements of grounded and sensorimotor theories of cognition with dynamic models of decision-making, which are up to now disconnected but address categorization from different and potentially complementary perspectives.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements 270108 (Goal-Leaders, supporting GP), and 231281 (EuCogII, which sponsored JCQ's visit to ISTC-CNR in Rome). It has also received funding from the French program "investissement d'avenir" managed by

the National Research Agency (ANR), from the European Union (Auvergne European Regional Development Funds) and from the "Rgion Auvergne" in the framework of the IMobS3 LabEx (ANR-10-LABX-16-01) (which sponsored GP's visit to Pascal Institute in Clermont-Ferrand).

References

Akrami, A., Liu, Y., Treves, A., Jagadeesh, B., Apr 2009. Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli. Cereb Cortex 19 (4), 760–776.

Amit, D. J., 1989. Modeling Brain Function, The world of Attractor Neural Network. Cambridge University Press.

Amit, D. J., Fusi, S., Yakovlev, V., Jul 1997. Paradigmatic working memory (attractor) cell in it cortex. Neural Comput 9 (5), 1071–1092.

Ballard, D., 1991. Animate vision. Artificial Intelligence 48 (1), 1–27.

Barca, L., Pezzulo, G., 2012. Unfolding visual lexical decision in time. PLoS ONE 7 (4), e35932.

Barsalou, L., 2008. Grounded cognition. Annual Review of Psychology 59, 617–645.

Barsalou, L., Wiemer-Hastings, K., 2005. Situating abstract concepts. In: Pecher, D., Zwaan, R. (Eds.), Grounding cognition: The role of perception and action in memory, language, and thought. Cambridge University Press, New York, pp. 129–163.

Barsalou, L. W., May 1983. Ad hoc categories. Mem Cognit 11 (3), 211–227.

Barsalou, L. W., 1999. Perceptual symbol systems. Behavioral and Brain Sciences 22, 577–600.

Barsalou, L. W., 2003. Situated simulation in the human conceptual system. Language and Cognitive Processes 18, 513–562.

Barsalou, L. W., May 2009. Simulation, situated conceptualization, and prediction. Philos Trans R Soc Lond B Biol Sci 364 (1521), 1281–1289.

Bassett, D. S., Gazzaniga, M. S., May 2011. Understanding complexity in the human brain. Trends Cogn Sci 15 (5), 200–209.

Bickhard, M. H., 1993. Representational content in humans and machines. Journal of Experimental and Theoretical Artificial Intelligence 5, 285–333.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J. D., Oct 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol Rev 113 (4), 700–765.

Chao, L. L., Martin, A., Oct 2000. Representation of manipulable man-made objects in the dorsal stream. Neuroimage 12 (4), 478–484.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., Shenoy, K. V., Nov 2010. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? Neuron 68 (3), 387–400.

Cisek, P., Jun 2012. Making decisions through a distributed consensus. Curr Opin Neurobiol.

Cisek, P., Kalaska, J. F., 2010. Neural mechanisms for interacting with a world full of action choices. Annu Rev Neurosci 33, 269–298.

Clifford, C. W., Beardsley, S. A., Vaina, L. M., 1999. The perception and discrimination of speed in complex motion. Vision Research 39.

Coles, M. G., Gratton, G., Bashore, T. R., Eriksen, C. W., Donchin, E., Oct 1985. A psychophysiological investigation of the continuous flow model of human information processing. J Exp Psychol Hum Percept Perform 11 (5), 529–553.

Desmurget, M., Grafton, S., 2000. Forward modeling allows feedback control for fast reaching movements. Trends Cogn. Sci. 4, 423–431.

Dindo, H., Zambuto, D., Pezzulo, G., 2011. Motor simulation via coupled internal models using sequential monte carlo. In: Proceedings of IJCAI 2011. pp. 2113–2119.

Fagioli, S., Hommel, B., Schubotz, R. I., 2007. Intentional control of attention: action planning primes action-related stimulus dimensions. Psychol Res 71 (1), 22–29.

Frintrop, S., Jensfelt, P., 2008. Attentional landmarks and active gaze control for visual slam. Robotics, IEEE Transactions on 24 (5), 1054–1065.

Friston, K., Apr 2005. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360 (1456), 815–836.

Friston, K., Adams, R. A., Perrinet, L., Breakspear, M., 2012. Perceptions as hypotheses: saccades as experiments. Front Psychol 3, 151.
URL `http://dx.doi.org/10.3389/fpsyg.2012.00151`

Geisler, W. S., Apr 2011. Contributions of ideal observer theory to vision research. Vision Res 51 (7), 771–781.

Gigliotta, O., Nolfi, S., 2008. On the coupling between agent internal and agent/ environmental dynamics: Development of spatial representations in evolving autonomous robots. Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems 16 (2-3), 148–165.

Gold, J., Shadlen, M., 2001. Neural computations that underlie decisions about sensory stimuli. Trends in Cognitive Sciences 5, 10–16.

Gold, J. I., Shadlen, M. N., 2007. The neural basis of decision making. Annu Rev Neurosci 30, 535–574.

Grush, R., Jun 2004. The emulation theory of representation: motor control, imagery, and perception. Behavioral and Brain Sciences 27 (3), 377–96.

Hayhoe, M., Ballard, D., Apr 2005. Eye movements in natural behavior. Trends Cogn Sci 9 (4), 188–194.

Hommel, B., Musseler, J., Aschersleben, G., Prinz, W., 2001. The theory of event coding (tec): a framework for perception and action planning. Behavioral and Brain Science 24(5), 849–78.

Hopfield, J., 1982. Neural networks and physical systems with emergent, collective computational properties. Proc. Natl. Acad. Sci. USA 79, 2554–2558.

Itti, L., Koch, C., Niebur, E., Nov 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (11), 1254–1259.

Jeannerod, M., 2001. Neural simulation of action: A unifying mechanism for motor cognition. NeuroImage 14, S103–S109.

Jeannerod, M., 2006. Motor Cognition. Oxford University Press.

Kawato, M., 1999. Internal models for motor control and trajectory planning. Current Opinion in Neurobiology 9, 718–27.

Kietzmann, T. C., Geuter, S., Koenig, P., 2011. Overt visual attention as a causal factor of perceptual awareness. PLoS One 6 (7), e22614.

Kilner, J., Paulignan, Y., Blakemore, S., 2003. An interference effect of observed biological movement on action. Current Biology 13, 522–525.

Kokinov, B. N., 1994. The context-sensitive cognitive architecture dual. In: Proceedings of Cogsci XVI. Lawrence Erlbaum Associates.

Krajbich, I., Armel, C., Rangel, A., Oct 2010. Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci 13 (10), 1292–1298.
URL http://dx.doi.org/10.1038/nn.2635

Kruschke, J. K., 1992. Alcove: An exemplar-based connectionist model of category learning. Psychological Review 99, 22–44.

Lamberts, K., Apr 2000. Information-accumulation theory of speeded categorization. Psychol Rev 107 (2), 227–260.

Lepora, N., Martinez, U., Prescott, T., 2013. Active touch for robust perception under position uncertainty. In: IEEE proceedings of ICRA.

Lowe, D. G., 1999. Object recognition from local scale-invariant features. In: Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, pp. 1150–1157.

Maye, A., Engel, A. K., 2011. A computational model of sensorimotor contingencies for object perception and control of behavior. In: IEEE International Conference on Robotics and Automation (ICRA 2011).

Medin, D. L., Schaffer, M. M., 1978. Context theory of classification learning. Psychological Review 85, 207238.

Mirolli, M., Ferrauto, T., Nolfi, S., 2010. Categorisation through evidence accumulation in an active vision system. Connection Science 22 (4), 331–354.

Miyashita, Y., Chang, H., 1988. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature 331 (68).

Nelson, J. D., Cottrell, G. W., 2007. A probabilistic model of eye movements in concept formation. Neurocomput. 70 (13-15), 2256–2272.

Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., Sejnowski, T. J., Jul 2010. Experience matters: information acquisition optimizes probability gain. Psychol Sci 21 (7), 960–969.

Nolfi, S., Marocco, D., 2002. Evolving robots able to visually discriminate between objects with different sizes. International Journal of Robotics and Automation 17 (4), 163–170.

Nosofsky, R. M., Palmeri, T. J., Apr 1997. An exemplar-based random walk model of speeded classification. Psychol Rev 104 (2), 266–300.

Olman, C., Kersten, D., 2004. Classification objects, ideal observers and generative models. Cognitive Science 28 (2), 227–239.

O'Regan, J., Noe, A., 2001. A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences 24 (5), 883–917.

Pezzulo, G., 2008. Coordinating with the future: the anticipatory nature of representation. Minds and Machines 18 (2), 179–225.

Pezzulo, G., 2011. Grounding procedural and declarative knowledge in sensorimotor anticipation. Mind and Language 26 (1), 78–114.

Pezzulo, G., 2013. Why do you fear the bogeyman? an embodied predictive coding model of perceptual inference. Cognitive, Affective, and Behavioral Neuroscience.

Pezzulo, G., Barsalou, L., Cangelosi, A., Fischer, M., McRae, K., Spivey, M., 2011. The mechanics of embodiment: A dialogue on embodiment and computational modeling. Frontiers in Psychology 2 (5), 1–21.

Pezzulo, G., Barsalou, L., Cangelosi, A., Fischer, M., McRae, K., Spivey, M., 2013. Computational grounded cognition: A new alliance between grounded cognition and computational modeling. Frontiers in Psychology.

Pezzulo, G., Calvi, G., 2007. Designing modular architectures in the framework akira. Multiagent and Grid Systems 3 (1), 65–86.

Pezzulo, G., Calvi, G., 2011. Computational explorations of perceptual symbol system theory. New Ideas in Psychology 29 (3), 275–297.

Pezzulo, G., Castelfranchi, C., 2009. Thinking as the control of imagination: a conceptual framework for goal-directed systems. Psychological Research 73 (4), 559–577.

Quinton, J. C., Catenacci Volpi, N., Barca, L., Pezzulo, G., 2013. The cat is on the mat. or is it a dog? dynamic competition in perceptual decision making. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems.

Quinton, J.-C., Girau, B., 2012. Spatiotemporal pattern discrimination using predictive dynamic neural fields. BMC Neuroscience 13 (Suppl.1), 1–2.

Rao, R. P., Ballard, D. H., January 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2 (1), 79–87.

Ratcliff, R., 1978. A theory of memory retrieval. Psychological Review 85, 59–108.

Ratcliff, R., Rouder, J. N., 1998. Modeling response times for two-choice decisions. Psychological Science 9 (5), 347–356.

Rehder, B., Hoffman, A. B., Aug 2005. Eyetracking and selective attention in category learning. Cogn Psychol 51 (1), 1–41.

Resulaj, A., Kiani, R., Wolpert, D. M., Shadlen, M. N., August 2009. Changes of mind in decision-making. Nature 461 (7261), 263–266.

Rizzolatti, G., Riggio, L., Dascola, I., Umilt, C., 1987. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. Neuropsychologia 25 (1A), 31–40.

Rojas, R., 1996. Neural Network: A Schematic Introduction. Springer-Verlag.

Rosch, E., 1975. Cognitive representations of semantic categories. Journal of Experimental Psychology: General 104, 192233.

Rothkopf, C. A., Ballard, D. H., Hayhoe, M. M., 2007. Task and context determine where you look. J Vis 7 (14), 16.1–1620.

Roy, D., 2005. Semiotic schemas: a framework for grounding language in action and perception. Artificial Intelligence 167 (1-2), 170–205.

Sakai, K., Miyashita, Y., Nov 1991. Neural organization for the long-term memory of paired associates. Nature 354 (6349), 152–155.

Salinas, E., Sejnowski, T. J., 2001. Gain modulation in the central nervous system: Where behavior, neurophysiology, and computation meet. The Neuroscientist 7 (5), 430–440.

Sanborn, A. N., Griffiths, T. L., Shiffrin, R. M., Mar 2010. Uncovering mental representations with markov chain monte carlo. Cogn Psychol 60 (2), 63–106.

Schoener, G., 2008. Dynamical systems approaches to cognition. In: Spencer, P., Thomas, M., McClelland, J. (Eds.), Toward a Unified Theory of Development: Connectionism and Dynamic Systems Theory Re-Considered. ford University Press.

Shadlen, M., Kiani, R., Hanks, T., Churchland, A., 2008. Neurobiology of decision making: An intentional framework. In: Engel, C., Singer, W. (Eds.), Better than conscious?: decision making, the human mind, and implications for institutions. The MIT Press.

Song, J.-H., Nakayama, K., Aug 2009. Hidden cognitive states revealed in choice reaching tasks. Trends Cogn Sci 13 (8), 360–366.

Spivey, M., 2007. The continuity of mind. Oxford University Press, USA.

Strauss, S., Heinke, D., 2012. A robotics-based approach to modeling of choice reaching experiments on visual attention. Front Psychol 3, 105.

Tipper, S., Howard, L., Houghton, G., 2000. Behavioural consequences of selection from population codes. In: Monsell, S., Driver, J. (Eds.), Attention and performance XVII. MIT Press, Boston. MIT Press, Cambridge, MA, pp. 223–246.

Tosoni, A., Galati, G., Romani, G. L., Corbetta, M., Dec 2008. Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions. Nat Neurosci 11 (12), 1446–1453. URL `http://dx.doi.org/10.1038/nn.2221`

Trabasso, T., Bower, G., 1968. Attention in learning: Theory and research. Wiley, New York.

Tuci, E., Massera, G., Nolfi, S., 2010. Active categorical perception of object shapes in a simulated anthropomorphic robotic arm. Transaction on Evolutionary Computation Journal 14 (6), 885–899.

Tucker, M., Ellis, R., 2004. Action priming by briefly presented objects. Acta Psychol. 116, 185–203.

Usher, M., McClelland, J. L., 2001. On the time course of perceptual choice: The leaky, competing accumulator model. Psychological Review 108(3), 550–592.

Wang, X.-J., Dec 2002. Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36 (5), 955–968.

Wolpert, D. M., Doya, K., Kawato, M., Mar 2003. A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci 358 (1431), 593–602.

Yarbus, A., 1967. Eye Movements and Vision. Plenum Press, New York.