

***Citation for the published version:***

Quinton, J. C., Catenacci Volpi, N., Barca, L., & Pezzulo, G. (2014). 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 and Humans, 44(5), 539-551. DOI: 10.1109/TSMC.2013.2279664

***Document Version:*** Accepted Version

***Link to the final published version available at the publisher:***

<https://doi.org/10.1109/TSMC.2013.2279664>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

***General rights***

Copyright© and Moral Rights for the publications made accessible on this site are retained by the individual authors and/or other copyright owners.

Please check the manuscript for details of any other licences that may have been applied and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights. You may not engage in further distribution of the material for any profitmaking activities or any commercial gain. You may freely distribute both the url (<http://uhra.herts.ac.uk/>) and the content of this paper for research or private study, educational, or not-for-profit purposes without prior permission or charge.

***Take down policy***

If you believe that this document breaches copyright please contact us providing details, any such items will be temporarily removed from the repository pending investigation.

***Enquiries***

Please contact University of Hertfordshire Research & Scholarly Communications for any enquiries at [rsc@herts.ac.uk](mailto:rsc@herts.ac.uk)

# The cat is on the mat. Or is it a dog?

## Dynamic competition in perceptual decision making

Jean Charles Quinton, Nicola Catenacci Volpi, Laura Barca, and Giovanni Pezzulo

**Abstract**—Recent neurobiological findings suggest that the brain solves perceptual decision-making tasks by means of a dynamic competition in which evidence is accumulated in favor of the alternatives. However, it is unclear if and how the same process applies in more complex, real-world tasks such as the categorization of ambiguous visual scenes and what elements are considered as evidence in this case. Furthermore, dynamic decision models typically consider evidence accumulation as a passive process disregarding the role of active perception strategies. In this article, we adopt the principles of dynamic competition and active vision for the realization of a biologically-motivated computational model, which we test in a visual categorization task. Furthermore, our system uses predictive power of the features as the main dimension for both evidence accumulation and the guidance of active vision. Comparison of human and synthetic data in a common experimental set-up suggests that the proposed model captures essential aspects of how the brain solves perceptual ambiguities in time. Our results point to the importance of the proposed principles of dynamic competition, parallel specification and selection of multiple alternatives, prediction, and active guidance of perceptual strategies for perceptual decision-making and the solution of perceptual ambiguities, and suggest that they could apply to both the simple perceptual decision problems studied in neuroscience and the more complex ones addressed by vision research.

**Index Terms**—Perceptual decision-making, dynamic models, prediction, active vision.

### I. INTRODUCTION

ACCORDING to an authoritative view in neuroscience, perceptual decision-making is a dynamic process in which alternatives (e.g., are these dots moving towards left or right?) compete over time [1], [2]. Perceptual ambiguities are solved by accumulating sensory evidence in favor of the hypotheses, up to a threshold. Initial decisions can be revised, too, when evidence is initially stronger for one hypothesis, and successively for another. This process is well captured by dynamic models of choice, such as drift-diffusion [3], neural

races [4], dynamic accumulators [5], and dynamic stochastic models [6], all essentially implementing statistical tests [7].

Dynamic models of choice have a good degree of correspondence with neural data. Indeed, numerous experiments have revealed that the activation of neurons in sensorimotor areas (e.g., the macaque LIP, lateral intraparietal cortex) “ramps” up in a way that is consistent with drift-diffusion models, and is highly predictive of overt response (e.g., overt eye movements to the left or right). This body of evidence has led to the proposal of a so-called *intentional framework* of perceptual decision-making [8], in which evidence accumulation is intimately related to action selection. Neuroimaging experiments in humans show that the same mechanisms might be in play besides simple perceptual decisions, such as for instance in the recognition of faces and buildings, and could support quite arbitrary mappings between stimuli and actions [9]. Recent studies indicate that a similar “ramping” mechanism could regulate more abstract decisions that are not tied to any effector-specific response [10]. Besides perceptual decisions, dynamic competition has been proposed as a key principle for the parallel specification and selection of multiple responses [11] and categorization [12], [13]. In brief, this corpus of evidence suggests the brain solves choice problems (of many or possibly all kinds) using dynamic competition between two or more hypotheses (or responses) maintained and updated in parallel.

Although the proposals reviewed so far tend to emphasize bottom-up sensory processes, a complementary view that is gaining prominence is that perceptual processing is inherently proactive and anticipatory [14], [15]. In a similar vein, it has been proposed that to solve perceptual uncertainties the brain adopts a generative approach to perceptual processing. In this *predictive coding* view, the brain builds a hierarchical, statistical model of the sensorium, and uses it to guide perceptual processing. Higher (cortical) layers represent increasingly more abstract object features (or even semantic information), and bias lower layers in a top-down manner by propagating *expectations*, which play the role of (Bayesian) *priors*. In turn, lower layers, which encode more fine-grained details of perceptual stimuli, provide bottom-up feedback in the form of prediction errors, which guide revisions of the perceptual hypotheses [16], [17], [18]. In this framework, choice is operated by minimizing prediction errors (generated by the competing hypotheses) rather than accumulating evidence in favor of the alternatives.

Hierarchical and generative approaches to visual processing are becoming popular in vision research, too (although discriminative methods are still widespread). In the last few

J-C. Quinton is with Clermont University, Blaise Pascal University, Pascal Institute, BP 10448, F-63000 Clermont-Ferrand / CNRS, UMR 6602, Pascal Institute, F-63171 Aubiere, France. e-mail: j-charles.quinton@univ-bpclermont.fr.

N. Catenacci Volpi is with IMT Institute for Advanced Studies, Piazza S. Ponziano 6 - 55100 Lucca, Italy. e-mail: nicola.catenacci@imtlucca.it.

L. Barca, is with the Institute of Cognitive Sciences and Technologies - National Research Council of Italy, Via S. Martino della Battaglia 44 - 00185 Rome, Italy. e-mail: laura.barca@istc.cnr.it.

G. Pezzulo is with the Institute of Computational Linguistics “Antonio Zampolli” - National Research Council of Italy, Via Giuseppe Moruzzi, 1 - 56124 Pisa, Italy and with Institute of Cognitive Sciences and Technologies - National Research Council of Italy, Via S. Martino della Battaglia 44 - 00185 Rome, Italy. e-mail: giovanni.pezzulo@istc.cnr.it.

Manuscript received ???, ???; revised ???, ???.

years, considerable progresses have been made towards the realization of robust and scalable, so-called “deep learning” architectures [19], [20]. Despite so, generative methods are generally considered hard to implement and to scale up to realistic situations. Possible solutions to this problem consist in extracting and using (hierarchies of) features, which permit recognizing equivalent object parts despite their different appearance [21], [22], combining generative and discriminative approaches, or incorporating human knowledge in the decision process.

A limitation of all these approaches is that they incorporate a passive model of information collection. On the contrary, the *active perception* view emphasizes that living organisms gather information is an active way by exerting control over their sensors and effectors [23]. As they are able to (partially) determine their next stimuli, they can use an active strategy to select what visual stimuli to attend and how to probe the visual scene. In this perspective, perceptual decision-making can be significantly affected by the strategy used to scan the visual scene. A recent study shows that the sequence of eye-movements affects recognition and categorization of ambiguous figures [24]. This result can be explained in terms of dynamic decision-making theories by noticing that, by performing different scans of the same visual scene, subjects accumulate more evidence that is consistent with one or the other alternative. This also implies that (active) sensing is part of the decision-making process, and not only a source of input. In keeping with this view, active vision schemes can be adopted to select information in a top-down manner, depending on the demands of the task at hand. There is indeed abundant evidence of top-down guidance of perception and attention strategies in naturalistic environments, and the anticipatory search of relevant information [25]. Recently, these insights have been incorporated in multiple-model architectures of reinforcement learning, which are able to select gaze locations depending on utility (or losses) associated to multiple goals [26], [27]. However, these architectures do not address problems of perceptual categorization.

Overall, despite the advances in neuroscience and artificial vision research described so far, knowledge remains scattered among different communities of neuroscientists and vision researchers. On the one hand, although there is consensus among neuroscientists that perceptual decision-making is a dynamic and competitive process, it is still unclear how neural mechanisms studied by neuroscientists in simple perceptual choices can scale up to more complex visual categorization tasks, and what elements are used as evidence for the alternatives. On the other hand, the advanced computational methods developed by vision researchers and the insights coming from the study of complex visual tasks (such as the importance of feature-based representation, predictive processes, and the active guidance of perception) are rarely studied in neuroscience experiments, and their neural substrate is incompletely known. In this article we combine these complementary aspects in a unique system, which we then test in a human categorization study.

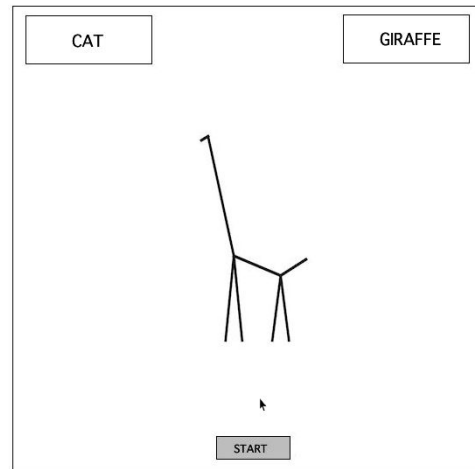


Fig. 1. Sketch of the set-up used in both human and simulated experiments. After pressing a START button, a visual stimulus is presented centrally. The task consists in classifying it by pressing (with the mouse) one of the two buttons (CAT or GIRAFFE). Measuring mouse kinematics permits to unfold the visual decision in time.

#### A. A dynamical approach to perceptual decision-making

We pursue a novel approach that combines aspects of dynamical systems, predictive coding, (modularized) feature-based schemes, and active vision systems described so far into an integrated computational architecture for perceptual categorization. Visual stimuli categorization is modeled as a dynamic process, in which alternative hypotheses (e.g., giraffe vs. dog) compete over time. Different from most bottom-up approaches, in the proposed model competition is solved by considering the prediction success of the alternative hypotheses. In brief (and with some simplifications, see later), the neural architecture that we present includes sets of feature predictors, specific for each category, which continuously “vote” for one of the alternatives; their votes are only counted if predictions are correct.

Not only prediction success is used for choice, but also for selecting the next gaze location of the system fovea. In this way, the currently leading hypothesis is also the one that steers an active vision process and influences the way the visual stimulus is probed. As it can influence choice, the active guidance of the fovea can be considered as a part of the decision-making process. The use of active perception mechanisms (guided by prediction success) along with a realistic implementation of action dynamics and stimuli processing distinguishes our proposal from related models of categorization that emphasize dynamic processing [12], competition between features of objects [13] or their predictive power [28].

To demonstrate the efficacy of our computational approach, we compare human and system performance in a visual categorization task. The set-up used in both human and simulated experiments is sketched in Fig. 1. Essentially, for both the human and the system, the task consists in classifying a (shortly presented) visual stimulus as belonging to one of two categories (e.g., dog vs. giraffe), by clicking one of two buttons with the mouse. In the experiments, we used visual stimuli (sketches of animal pictures) belonging to four categories (cat,

dog, giraffe, and horse) and having two different levels of ambiguity (low or high). For instance, in a Giraffe vs. Dog classification, stimuli can be of four different kinds: “prototypical” Giraffes, “prototypical” Dogs, or figures obtained by interpolating the two and conserving more elements from the former or the latter, respectively. In the first two cases, stimuli had low ambiguity; in the last two cases, stimuli had high ambiguity.

To visualize and measure the dynamics of choice in the human experiment, or their “decision trajectory”, we measured the continuous mouse movements they made during the task. This methodology is widely adopted to study the dynamics of decision making and how they change as a function of uncertainty, see e.g., [29], [30]<sup>1</sup>. We hypothesized that, in the presence of ambiguous figures, human subjects would have made more classification errors. Furthermore, and more significantly for our study, we hypothesized that the trajectories of their mouse movements would have been less straight, as if they were more “attracted” by the other category (as compared to trials with non-ambiguous figures). This hypothesis is consistent with the view that perceptual uncertainties are solved by a dynamic competition process, in which interpretations are biased by evidence collected and prediction errors, and in which initial hypotheses can be revised during the processing.

Not only the continuous measurement of subjects’ mouse trajectories is important per se, but it also allows comparing human and system performance. To this aim, we tested our computational system in the same task as the human subjects. For each trial, we simulated mouse trajectories by counting (step-by-step) the “votes” in favor of each alternative, and by generating a movement vector pointing towards the currently preferred alternative. Choices that were certain from the beginning resulted in straight trajectories towards the winning alternative, and uncertain choices generated curve trajectories and “changes of mind” during the task.

Below we introduce in more detail the human (section II) and simulated experiments (section III) we performed.

## II. METHODS: THE HUMAN EXPERIMENT

### A. Ethics Statement

The procedure of the experiment has been approved by the Institute of Cognitive Sciences and Technologies of the National Research Council, ISTC-CNR of Rome, Italy. All the participants gave their informed consents.

### B. Participants

Eighteen participants took part in the experiment. Age ranged from 25 to 63 years. All participants were Italian native speakers, highly educated (university students, people with master degree or young researchers), and with normal or corrected to normal vision.

<sup>1</sup>Note that this methodology is based on the assumption that decision-making is *not* completed when the action starts, but is a continuous process [31]. Traditional models of decision-making do not readily accommodate this assumption, although they can be extended to do so [32].

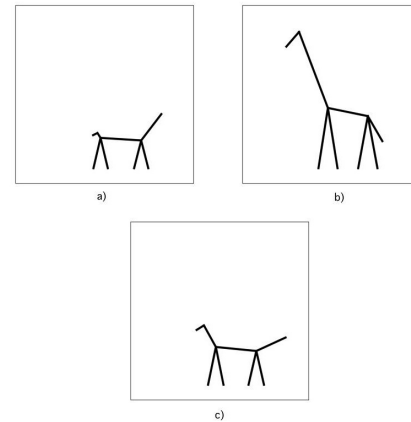


Fig. 2. a) Example of a stylized Cat. b) Example of stylized Giraffe. c) Example of an ambiguous Cat that is also 25% a Giraffe.

### C. Materials

A list of ambiguous and unambiguous figures has been used, for a total of 96 experimental stimuli. The 48 unambiguous figures represent stylized animals belonging uniformly to four possible categories: Cat, Dog, Giraffe and Horse. The stimuli used in this experiment were originally developed by [33]. The category prototypes were determined by [34]: one thousand stimuli were initially generated by a software by considering nine parameters (e.g., head angle, tail length), and then categorized by eight participants. The average of responses among these trials produced eight prototype figures (one for each participants) for each category.

The 48 ambiguous figures used in our study were generated by randomly choosing two prototypes belonging to different categories (e.g., Cat and Giraffe), and then applying a morphing procedure, which varied their nine parameters along a continuum between the two categories (e.g., between the head angle mean of Cat and Giraffe). This procedure permitted to produce figures that are “intermediate” between two categories, using a morphing coefficient  $\mu$  in  $[0, 1]$ . Specifically, we selected figures belonging for the 75% to one of the two categories, and 25% to the other (e.g., an ambiguous Cat that is also 25% Giraffe). Examples of unambiguous and ambiguous figures of stylized animal are shown in Fig. 2(a,b) and Fig. 2(c), respectively.

### D. Procedure

At the beginning of each trial, participants clicked on the /START/ button located bottom-centre of the PC screen (see Fig. 1). Then stimuli appeared and participants had 800ms to make their response (i.e., to move the mouse cursor and click on one of the two response buttons), otherwise a /TIME OUT/ message appeared. Response buttons were labelled with words standing for categories (e.g., /CAT/). One button was always associated with the correct response category, while the other was either a random category (for unambiguous stimuli) or the other category considered in the morphing procedure described so far (in the case of ambiguous stimuli). In other

words, ambiguous stimuli were always used in decisions concerning the two morphed categories. Words were presented in ARIAL font, upper case, black on a white background, located respectively top-left or top-right of the PC screen, on the basis of a random choice. Also correspondence between stimulus type and button were varied randomly across trials and participants. In case of errors, a feedback message (red cross) appeared after the response.

During subject responses, categorization errors and the  $x$  and  $y$  coordinates of mouse trajectories were recorded automatically (sampling rate of approximately 70Hz) using *MouseTracker*. Such software package was used to record, process, and analyze mouse movements [35]. Before starting with experimental data acquisition, participants performed a practice session with 10 items to familiarize with the procedure. The 96 experimental stimuli were presented in two blocks of 48 items each. The order of stimuli within blocks and the order of blocks presentation were randomized.

### E. Results

1) *Data Processing*: Error rate and mouse trajectories were analyzed under two conditions: in the presence of ambiguous and unambiguous stimuli. In the trajectory analysis we discarded wrong responses (i.e., when the subject selected the inappropriate stimulus category) which were the 19% of total data. This high error rate is explained by the task speed constraint and the ambiguity of half the stimuli.

We used the Linear Mixed-Effects Model (LMM) to study the effect of ambiguity on the response variables. Stimulus ambiguity was considered as a Fixed-effect factor and Items and Subjects were considered as Random-effects factors. Among the two conditions the unambiguous one was taken as default level of comparison.

For each dependent variable (i.e., trajectory's parameters and accuracy) an independent analysis was conducted. Analyses were run with the *lm4* package for R [36], where  $p$  values were estimated by using Markov chain Monte Carlo simulations [37].

2) *Accuracy Rate Analysis*: For each condition, mean accuracy rate was computed across all participants and trials. As expected, participants were less accurate in categorizing ambiguous stimuli ( $M = 0.25, SD = 0.43$ ) than unambiguous stimuli ( $M = 0.14, SD = 0.35$ ). Such difference was statistically significant as shown by mixed-effects models on accuracy rate ( $\beta = 0.102, p_{MCMC} < 0.008$ ).

3) *Trajectories Analysis*: In Fig. 3, the mean trajectories of the two conditions across all trials are shown. The mean trajectory for the ambiguous condition has a pronounced curvature, whereas the mean trajectory for the unambiguous condition is closer to the ideal straight line of response, meaning that trajectories for the ambiguous condition were more attracted to the opposite alternatives. Curvature of the trajectory signals that uncertainty of choice is present during the time course of the decision.

To analyze formally the structure of trajectories, we considered their Area Under the Curve (AUC), which is a measure of spatial attraction toward the unselected alternative. It is

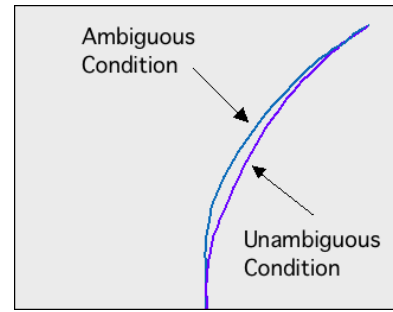


Fig. 3. Mean trajectories in the two conditions. Note that they were “remapped”, as if all responses were made to the right.

calculated as the geometric area between the actual trajectory and the idealized straight line trajectory from the /START/ button toward the selected response. Again, the mean value and standard deviation for the ambiguous stimuli ( $M = 0.92, SD = 1.59$ ) were higher than those in the unambiguous condition ( $M = 0.74, SD = 1.44$ ).

AUC was analyzed using the Linear Mixed-Effects model with stimulus ambiguity as Fixed-effects factor and Subjects and Items as two Random-effects factors. The unambiguous condition was taken as reference condition. Positive contrast coefficient for ambiguous stimuli ( $\beta = 0.17, p_{MCMC} < 0.05$ ) shows that average AUC was significantly higher for unambiguous stimuli.

### F. Discussion

Consistent with our hypothesis, uncertainty in visual stimuli affects participants' performance, with higher number of errors and trajectory curvature for ambiguous items. We focus on the analysis of movement trajectories, which is more informative than errors of the dynamics of choice. The fact that trajectories are more attracted by the unselected category in ambiguous conditions (compared to unambiguous conditions) is consistent with previous categorization studies [38] and can be explained within a dynamical system framework, in which multiple options are computed in parallel and compete over time [31]. In our task, the unselected category is able to function (quite literally) as an “attractor” of the choice (and the mouse trajectory). In the presence of uncertainty, the attractor is stronger and exerts significant influence over the choice. The competition between alternatives is not solved prior to (overt) action onset, but continues during the task, as the partial results of the competition continuously flow into motor movements; this can explain the curvature of the (mouse) trajectories and the “changes of mind” as resulting from the dynamics and uncertainty during the choice. This pattern of results is not unique of our task, but has been consistently reported in several studies including lexical decision, numerical decision, and objects categorization [39], [29], [30]. To explain these findings, several researchers have developed dynamical models that incorporate competitive processes. For example, [40] implemented a dynamic competitive model of choice that successfully explains the results of a “phonological competitor” task [41] (but see [42] for an alternative explanation that consider perception as a serial rather than parallel process).

Eye-tracking studies in perceptual tasks (e.g., spoken word with phonological competitors [43], [44]) also support the idea of a dynamic competition by showing that a proportion of eye movements is directed to the unselected choice when it is phonologically similar to the correct one.

We consider this dynamical view of choice, which is largely consistent with neurophysiological evidence (collected using simpler set-ups, though), to be the key of our theoretical and computational proposal, and we argue that it applies to perceptual decisions at large. Below we introduce a computational model that uses dynamic competition at its core for dealing with perceptual ambiguities in the same task as described before. Not only our model incorporates the aforementioned aspects of dynamic competition and continuous flow of information, but it also has several distinguishing features. First, it considers prediction success as a source of evidence, thus highlighting the importance of predictive processes in the regulation of the competition between the competing choices. Prediction is used in combination with a feature-based representation of the stimuli, which permits to address perceptual categorization tasks that are significantly more complex than those adopted in the neuroscience literature. Second, it explicitly models the eye movements used for collecting evidence, using an active perception process that samples informative features from the stimulus. Third, it explicitly models the response dynamics (mouse movements).

### III. METHODS: THE COMPUTATIONAL MODEL

The computational model presented here implements categorization as a competition between two (or more) different sets of feature predictors, each associated with a prototype or figure, assumed to be known by the agent (see Fig. 4). To allow a fair comparison, the artificial system interacts with the exact same MouseTracker setup used for the human experiment (but we replaced words with stick figures). The purpose of this computational experiment is to show how dynamic decision-making can emerge from pure sensorimotor interaction, where distributed predictors continuously compete for action, thus not directly relying on more abstract concepts.

Two Java applets are provided in the supplementary material: the former demonstrates how stick figures were generated (*StickFigureGenerator.jar*), and the latter permits the user (or the artificial controller) to perform the task and see its results (*StickFigureDecision.jar*).

#### A. Input features

The input provided to the system is an equivalent of the view of the display for the human. It includes the current position of the mouse pointer, general information about the trial (targets categories, current feedback or signal provided to the user) and a small foveated area that can be freely moved around the stimulus. The system has only partial knowledge of the figure at any time, and has to actively explore it by generating saccades.

As our experiment focuses on the fast online recognition and disambiguation of figures, and the artificial controller does not benefit from the highly robust human visual system,

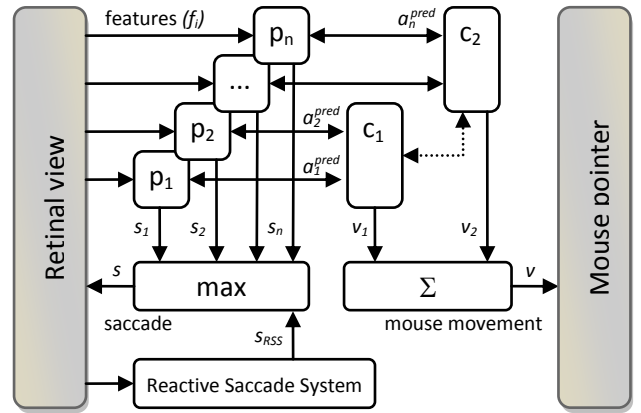


Fig. 4. Global design of the architecture. Plain arrows correspond to excitatory connections while the dashed arrow corresponds to the reciprocal inhibition between the categories. Whereas the saccade to be performed on the stick figure is chosen by the best predictor (or by a reactive saccade system if the predictors activity is too low), the mouse movement is a weighted sum of the vectors proposed for reaching the targets.

preprocessing the visual input is necessary to rival human performance (without a complex learning phase). To this aim, we adopted a feature-based representation of the stimuli. To select the right features (i.e. leading to feasible but not trivial discrimination between the stimuli) we computed neuro-inspired saliency maps of the stick figures [45]. Coherent with results on the use of coarse scales to rapidly categorize complex stimuli [46], a strong response could always be observed for contrast and orientation detectors near joints and would be sufficient for the task.

To lower the dimensionality of the input and make the model easier to interpret, only a set of feature points ( $f_i$ ) corresponding to the visible joints of the stick figure are retained for each fixation. In addition to its coordinates within the retinal image ( $u, v$ ), each feature is described by a vector synthesizing the oriented Gabor filters responses away from the joint ( $o_j$ ) $_{j \in [1, M]}$ . In practice, overlapping Gaussian tuning curves are used to directly generate the vector from the stick figure description (see Eq. 1). This method converts an arbitrary set of orientations into a fixed number of correlated activities (see Fig. 5). Notice that there is no way for the system to easily discriminate between the front and back legs without exploring the stimulus by generating saccades.

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

where  $\theta_j = -\pi + j\pi/M$  (with  $M$  fixed to 16 in the experiments), and  $\rho_l$  is the set of angles formed by the sticks starting from the considered joint.

Features are thus fully described by  $f_i = (u, v, o_1, \dots, o_M)$  and two features ( $f_1, f_2$ ) can be compared according to a similarity measure defined as:

$$\sigma(f_1, f_2) = 1 - e^{-\frac{\|f_2 - f_1\|^2}{\sigma^2}} \quad (2)$$

where  $\|\cdot\|$  is a norm in  $\mathbb{R}^{M+2}$  with an adequate weighting of the various dimensions involved.

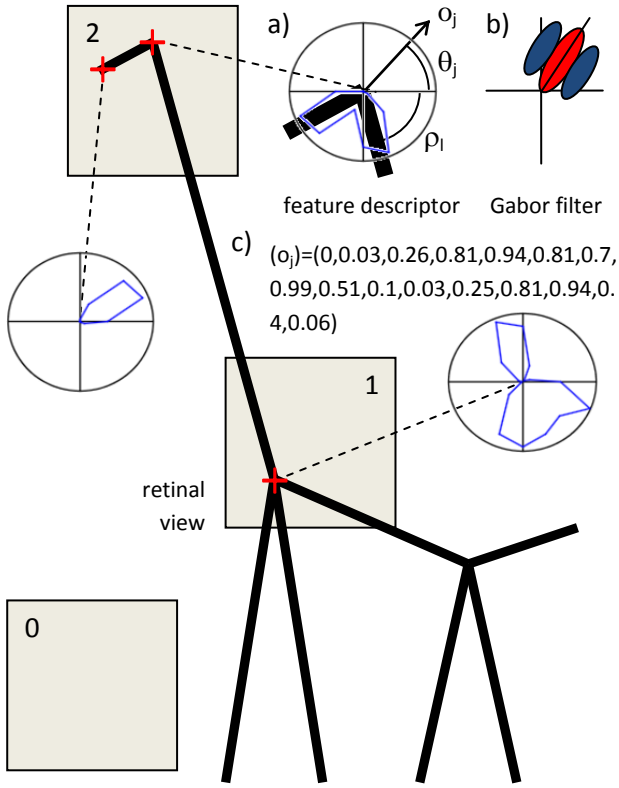


Fig. 5. For 3 arbitrary fixations on the stick figure, a variable number of feature points are extracted (*crosses*). a) The feature point descriptor is computed based on  $M$  different orientation detectors (*polar diagram*). b) Equivalent Gabor filter for the orientation detector corresponding to an angle  $\theta_j$ . c) Vector representation of the descriptor for the feature extracted from fixation 1.

### B. Output commands

The system controls both saccades and mouse movements. Cartesian coordinates are used to move the retina over the image, but there is no need for the sensations and commands to share the same coordinate system. Efferent signals remain separated from proprioceptive feedback, and their coupling is only achieved through the use of predictors at the core of the system (see below). Saccades are considered instantaneous, thus approximating the human system, where the fixation/saccade temporal ratio is high, and inhibition is often posited to occur during movement.

The mouse pointer controller also uses Cartesian coordinates in the screen frame of reference. A speed vector  $v$  is provided at all times by the predictive part of the architecture, and is integrated through an Euler scheme. Combined with a fixed speed limit, the system generates smooth trajectories on the screen with semi-realistic dynamics.

### C. Predictive representations

The core of the architecture is composed of a set of predictors ( $p_k$ ), each anticipating a feature  $f_k^{tgt}$  to be observed after performing a saccade  $s_k = (\delta x_k, \delta y_k)$  from an initial context where the feature  $f_k^{src}$  was present. Predictors are also specific to a stick figure and are thus also associated to the mouse

reaching movement towards the corresponding target category  $c_{\{1,2\}}$ . Such predictors introduce normativity at the core of the representations by simply suggesting the potentiality of observing some feature after performing an action, potentiality that can only be confirmed through interaction. This kind of models has proponents from theoretical [47], [48], [15], [49], experimental [50], and computational approaches [51], especially in the domain of neo-Piagetian and sensorimotor perspectives of cognition.

Each predictor continuously tries to assimilate the interactions the agent engages in, by updating a set of associated activities defined by:

$$\begin{aligned}
 a_k^{prop} &= a_k^{reac} * \sigma(s_k, \bar{s}) \\
 a_k^{pred} &= a_k^{prop} \times (\max_i \{\sigma(f_i, \overline{f_k^{tgt}})\} - \beta) \\
 a_k^{reac} &= \max_i \{\sigma(f_i, f_k^{src})\} \\
 a_k^{inhib} &= \max((1 - \alpha) a_k^{inhib}, a_k^{prop}) \\
 a_k &= a_k^{reac} - \gamma_2 * a_k^{inhib} + \gamma_1 * c_{\{1,2\}}
 \end{aligned} \tag{3}$$

where all overlined variables integrate the actual system dynamics, by compensating for induced shifts in commands and sensations when another predictor has been selected to control the retina.

$a_k^{prop}$  corresponds to the proprioceptive feedback determining if the predictor action is similar to the one previously selected.  $a_k^{pred}$  then evaluates the satisfaction of the expected consequences for this predictor, only if this predictor was compatible with the previous action taken. It is indeed possible to get a perfect match for features when non matching actions are performed if the stick figure belongs to another category.

$a_k^{reac}$  evaluates how much the predictor context matches the current situation, thus limiting the action selection mechanism to at least potentially adequate predictors.

$a_k^{inhib}$  is an inhibition of return term, preventing the system to keep checking the same successful predictions again and again. This contributes to implement a basic mechanism of *informativeness*, as there is little information to be obtained from the same fixation points on static stimuli.

Finally  $a_k$  combines these activities to determine if this predictor should be selected for controlling the saccade system. The  $c_{\{1,2\}}$  term biases the selection of predictors for categories that have already received support from the predictors, leading to an increased stability in the decision making process.

We assume that a set of adequate predictors has already been learned for each prototype, at least sufficient to discriminate between the proposed figures. This is coherent with the ability of humans to consistently associate a figure with an abstract concept such as an animal species or category. In accordance with the prototype theory of categorization [52] and the notion of conceptual spaces [53], the prototype has been chosen as the average figure in the multidimensional space of joint angles and stick lengths. However, we do not address here the debate between the different views on classification, and we could as well consider an arbitrary set of examples within one category while using the same similarity measure [54]. By simply sharing the category bias, basins of attraction corresponding to individual examples at the vision level can be merged into a single non convex representation of a category at the reaching level, whereas a prototype-based version implies convexity.

#### D. Action selection

The simplified algorithm below presents the global loop in which the perception and decision-making processes are performed. All predictors are updated based on the dynamics of the MouseTracker system, under the various task constraints and in response to the artificial agent's commands, as follows:

1. **for all** predictors ( $p_k$ )
2. Update  $a_k^{prop}$ ,  $a_k^{pred}$ ,  $a_k^{reac}$ ,  $a_k^{inhib}$ ,  $a_k$
3. **end**
4. Select the top  $a_k$  predictor to control the saccade
5. Move the retina to a new fixation point ( $x, y$ )
6. **for all** categories ( $c_i$ )
7. Update  $c_i$  with the best associated  $a_k^{pred}$
8. **end**
9. Make the categories ( $c_i$ ) compete for action
10. Move the mouse using the resulting speed vector

In addition to the predictors, a reactive saccade system is implemented to avoid the controller to saccade away from the stimulus. This might happen in the initial build-up phase of the predictors activity or if the stimuli is too dissimilar from any known category. In both cases, the system might select an inadequate predictor, that will lead to an area without any visible feature, thus preventing further recognition. In the same vein as the subsumption architecture [55], the reactive system can take control over the predictive system, inhibiting its output to lead the retina back to the closest salient point. A more dynamic way of presenting this mechanism is to consider competition between a relatively general yet static reactive process and several object-specific acquired predictive processes. The two mechanisms are complementary, reactivity being needed to bootstrap predictions and to cope with errors, while prediction is necessary for selective attention in active vision. In the end, the saccade to be performed is thus determined by:

$$s = \begin{cases} s_{argmax_k\{a_k\}} & \text{if } \max_k\{a_k\} > a_{RSS} \\ s_{RSS} & \text{otherwise} \end{cases} \quad (4)$$

where  $s_{RSS}$  and  $a_{RSS}$  respectively are the saccade and fixed activity of the reactive saccade system.

The control of the mouse movements is then adjusted in order to reflect the new information accumulated by the agent. For this purpose, the activities associated with each category are updated by accumulating evidence from the predictors, following equation 5.

$$\begin{aligned} \bar{c}_i &= \lambda c_i + \max_{k \in P_i} \{a_k^{pred}\} \\ c_i &= \frac{\bar{c}_i}{\sum_i \bar{c}_i} \end{aligned} \quad (5)$$

where  $P_i$  is the subset of predictors associated with target category  $c_i$  and  $\lambda$  an inertia coefficient, so as to limit the effect of isolated match or mismatch of expected consequences. As a side effect,  $\lambda$  also avoids abrupt changes in the motor commands, thus compensating for the absence of a more

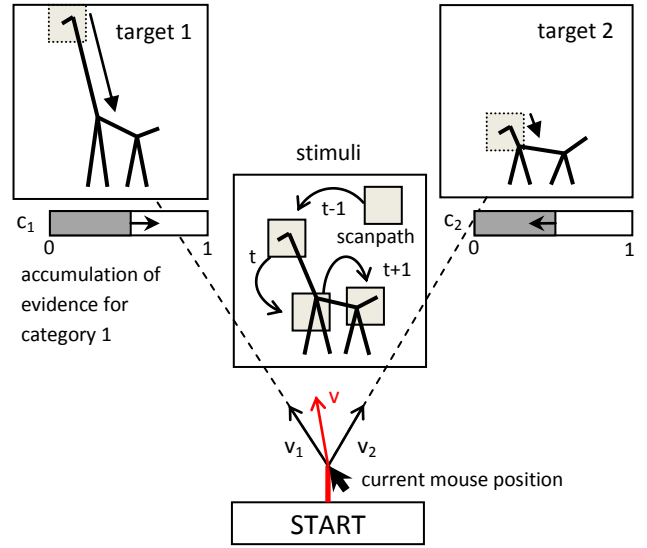


Fig. 6. Description of the predictive dynamics. Given that previous saccades before time  $t$  led to almost no discrimination between the prototypes, a fixation on the head of the figure will mainly activate the predictors associated with testing the neck length for both prototypes (i.e. saccading to the next salient feature at the shoulder). Once a saccade in this direction is performed, the system can measure the similarity of the observed features with expected consequences of the movement, here giving an advantage to the left prototype. If confirmed by further predictions, the resulting bias will allow the trajectory to bifurcate towards one of the target and the decision to be made.

realistic motor apparatus in the simulations. The normalization process used to obtain  $c_i$  from  $\bar{c}_i$  guarantees a bounded activity in  $[0, 1]$  and puts into competition the two categories. As our set-up only considers two categories for each trial, it does not require using more complex mechanisms of competition using reciprocal inhibition; however, the proposed model can be easily extended to incorporate such mechanisms. Indeed, attentional capabilities and increased robustness are required when the set of potentialities offered to the system increases. Committing to neuro-inspired models, dynamic neural field models allow such properties to emerge and can be applied to the sensory signals and categories [56] or directly to the predictors by using a high dimensional implementation [57].

Finally, the command sent for moving the pointer is a linear combination of the normalized vectors  $v_{\{1,2\}}$  aiming at the two visible targets from the current mouse position (see Fig. 6):

$$v = \sum_i c_i \times v_i \quad (6)$$

#### E. Results

1) *Trajectories Analysis:* Using the same protocol as in the human experiment, i.e. 18 artificial participants and 96 trials, the artificial system generates trajectories qualitatively similar to humans. Mean values and standard deviations obtained for the AUC measure are ( $M = 0.46, SD = 0.27$ ) for the ambiguous and ( $M = 0.30, SD = 0.10$ ) for the unambiguous stimuli. Using a paired T-test between the two conditions, the average AUC was significantly higher for ambiguous stimuli ( $p_{T-test} < 0.0001$ ). We also find significant but weaker results



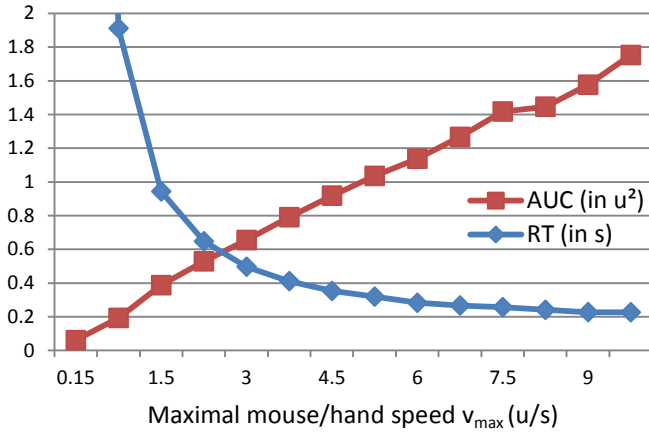


Fig. 7. Area Under Curve (*AUC*) and time to target (*RT*) as a function of the speed constraint, averaged over all prototypes, morphing coefficients and simulated participants. The target reaching time is limited by the amount of information gathered during each saccade, instead of following an inverse function of speed (e.g. 0.226 instead of 0.154 seconds for a speed of 9.75 screen units per second). This is also reflected in the *AUC*, that almost linearly increases with the speed constraint.

on reaction times (*RT*), underlying the interest of a finer analysis on the trajectory ( $M = 0.93$ ,  $SD = 0.02$  in unambiguous vs  $M = 0.96$ ,  $SD = 0.08$  in ambiguous condition). Variability in the computational model comes from the initial fixation point, which can make a huge difference in the bifurcation dynamics involved, due the complex system nature of the predictor-based controller. It is however much lower compared to the human data, as the simulated participants have a much lower intra-group variability.

The computational model we presented permits to run additional synthetic experiments by manipulating the system parameters. By increasing the maximal speed of mouse movements, and increasing the time constraint to reach a decision imposed to the system, higher standard deviation and contrast coefficient are obtained for the *AUC* distribution. Indeed, within a fixed amount of time, the same information will be gathered by the system through saccades, but a larger mouse movement will occur. Additionally, any error in the best predictor and saccade selection will lead to greater deviations from the ideal straight trajectory, as shown on Fig. 7 (all locations, speeds and *AUCs* are respectively expressed in MouseTracker units  $u$ , units per second, and squared units). On the contrary, when totally alleviating the time constraints, the average *AUC* is drastically decreased and there is no significant difference between the ambiguous and unambiguous conditions anymore. The dynamics of the decision process is then hidden in the first few milliseconds of the reaching movement. In a similar vein, human experiments have also shown that time constraints must be severe for other phenomena such as assimilation effects to be experimentally observed [58].

Using our computational system it is possible to increase the difficulty of the task by using highly ambiguous stimuli, up to a morphing coefficient  $\mu$  of 0.5 (this manipulation would be hard to do in human experiments, as in the presence of

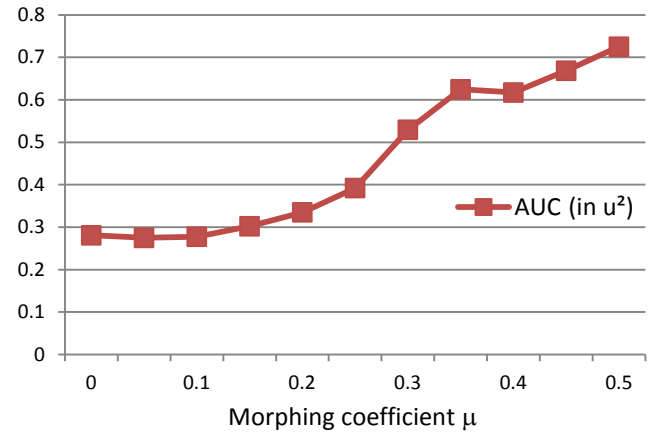


Fig. 8. Area Under Curve (*AUC*) as a function of the morphing coefficient  $\mu$  used for generating the stimuli, averaged over all prototypes and simulated participants. The *AUC* roughly follows a power function of the ambiguity, showing that coming to a definite decision becomes progressively harder when ambiguity increases, thus reflecting the non-linear system of equations governing the competition process between the two target prototypes.

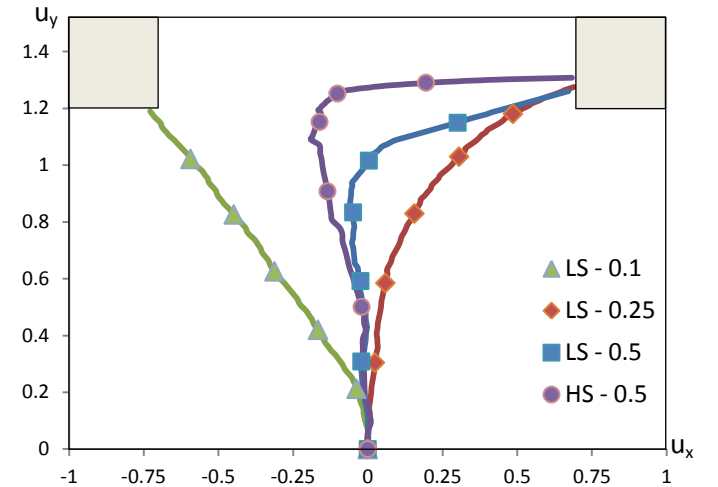


Fig. 9. Mouse trajectories generated by the artificial agent under various conditions. At low speed (*LS*), 3 representative trajectories are provided for a morphing factor in  $\{0.1, 0.25, 0.5\}$ , with increasing deviation from the straight trajectory. For high speed (*HS*) and high ambiguity (morphing coefficient of 0.5), the late change in decision during the reaching movement is amplified compared to the *LS - 0.5* condition.

stimuli that are too ambiguous subjects might adopt higher-level strategies, such as deliberately responding in a random manner). As in this system the decision process emerges from complex dynamic interactions between predictors, response accuracy does not degrade linearly (see Fig. 8). Nevertheless, for a morphing coefficient varying in  $[0.0, 0.5]$ , the *AUC* increases from 0.281 to 0.725. The values obtained for 0.0 and 0.25 are consistent with the above results on another set of data when also respectively considering 1.0 and 0.75 coefficients, showing that the attraction dynamics towards the prototypes is almost symmetric.

2) *Qualitative Analysis*: These quantitative results are explained by the non linearities and late changes in decisions that can be observed on individual reaching trajectories (see

Fig. 9). In turn, these phenomena result from the inertia of the predictors and the equal average amount of evidence accumulated for each prototype in the most ambiguous cases.

The non linear dynamics of the predictors activity comes from the feedback involved in the equations and through interaction with the MouseTracker environment. This feedback is modulated by the context, but even its nature (positive or negative) is determined by the satisfaction of the predictions. Predictors belonging to the same category are linked, so they benefit from the bias  $\gamma_1 * c_{\{1,2\}}$  in Eq. 3. At the same time, they influence each other through the environment, as an action triggered by one predictor can create a favorable context to trigger the activity of another predictor. Therefore, the movements of both the mouse and the eye have an influence on the overall dynamics. Reciprocally, the satisfaction of prediction term  $a_k^{pred}$  facilitates rapid bifurcations, as it contributes to saccade selection through the category bias, and determines mouse movements over time.

Predictors voting for the same category reinforce each other and steer the mouse movements towards their associated target. At the same time, they compete for controlling the saccades. At the category level, they discriminate among figures; individually, they discriminate among different parts of a figure. Overall, the competition dynamics emerges from a set of intertwined factors, which we list below (for simplicity, we only consider two arbitrary predictors):

- If they apply to different contexts ( $f_1^{src} \neq f_2^{src}$ ), the discrimination will occur at the sensory level, either because they are associated with different parts within the figures (head vs tail for instance) or because the stimulus highly differs from the prototype (e.g. different neck orientations).
- If they propose different actions ( $s_1 \neq s_2$ ), the one associated with the performed saccade will have better chances to see its expected consequences confirmed, as features must fall within the visual field. If the context and expected consequences are similar enough ( $f_1^{src} \simeq f_2^{src}$  and  $f_1^{tgt} \simeq f_2^{tgt}$ ), the proprioceptive feedback alone can lead to discrimination (e.g. short vs long neck).
- If only the expected consequences differ ( $f_1^{tgt} \neq f_2^{tgt}$ ), selecting the common action will lead to discrimination, either between prototypes or between parts of the figures (e.g. front vs back legs that can only be distinguished at the hip or shoulder level).
- If two predictors are almost identical, selecting them does not bring additional information for categorization, as they correspond to common parts of stick figures.

In the end, and due to competition between categories, only the relative distance to each prototype in the sensorimotor space matters for the final decision dynamics. This is true up to the discriminative power of the sensory apparatus, here limited by the standard deviation and norm characteristics in the similarity measure  $\sigma$  (see Eq. 2). However, this relative distance to the prototypes can only be ascertained statistically; by visually interacting with the stimulus, and under strict time constraints, this process leads to the complex dynamics observed.

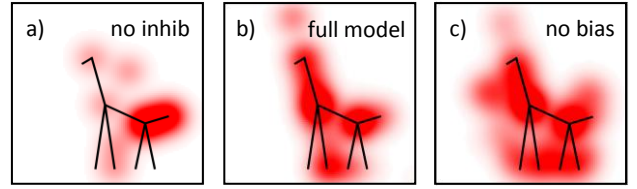


Fig. 10. Density of fixations at the end of a trial, in three cases: *a)* without the inhibition of return, *b)* full model, *c)* without the top-down bias. *a)* Without the simple embedded informativeness mechanism, the system has higher probability to converge on limit cycles of saccades where no decision can be made. *c)* When saccades are selected based on reactive mechanisms, the system tends to focus less on adequate features.

## F. Model decomposition

1) *Prediction*: To highlight the usefulness of prediction with ambiguous stimuli in the context of this study, we compared the proposed system with a reactive controller. In the reactive controller, observed features are passively matched to all features associated with each prototype figure to calculate  $a_k^{reac}$  (instead, the full system uses a mechanism for matching expected and observed features and calculating  $a_k^{pred}$ ).

Like the full system, the reactive controller produces significant differences for the AUC ( $M = 1.99, SD = 0.10$  for the unambiguous vs  $M = 2.13, SD = 0.08$  for the ambiguous condition,  $p_{T-test} < 0.0001$ ), and the RT ( $M = 5.19, SD = 2.05$  to  $M = 9.39, SD = 5.03$ ,  $p_{T-test} < 0.001$ ). The lack of the predictive component, however, leads to a significant increase in the mean values of both the AUC and RT, reflecting the difficulty and longer time required to reach a decision.

Outliers had to be removed for these quantitative results to be meaningful (RT above 3  $SD$  from the mean), even though a large standard deviation can still be observed for the reaction times in the ambiguous condition. These outliers provide an insight on the type of configurations that lead to non discriminant fixation patterns. Although the associated target prototypes may visually seem quite different, the set of features associated with the figure joints overlap. Fixations alone may be no more sufficient for the discrimination, and the organization of the joints should then be exploited through saccades.

Moreover, and following this qualitative line of reasoning, the non-predictive system also loses access to most information on the stick lengths since they are encoded in the saccades amplitude. This is experimentally confirmed by considering two categories where all features are identical, but for the length of a single stick (for instance a short-necked vs long-necked giraffe). Although the pattern of visual exploration still relies on a simple inhibition of return mechanism, thus saccading back and forth over most of the uninformative elements of the figure, the model without prediction cannot differentiate the two figures.

2) *Top-down modulation*: Reciprocal to the bottom-up influence of the prediction on decision making is the bias from the abstract category activity ( $c_{\{1,2\}}$ ) on the active vision process and the emergent saccade pattern. Nullifying  $\gamma_1$  removes the associated modulation of the predictors activity, and allows us to analyze the dynamics without any top-down

feedback.

This change in the equations is of course reflected in a loss of performance for the system, but its effect can be more easily interpreted on heatmaps representing the density of fixations on the stimulus (compare Fig.10b and 10c). By boosting predictors associated with the current most probable categories, the system limits the number of fixations away from the figure and thus actively improves the informativeness of the fixations and speeds up the convergence on a decision.

The fixation heatmaps presented here and the behavioral measures of the reaching dynamics introduced before do not aim at precisely mapping human data, but at reproducing qualitatively their pattern. Although this would have been possible, we did not tune the parameters of the system (listed in Table I) to exactly reproduce human data, because parameter values lying within an acceptable range (e.g. not making the whole figure visible within the field of view) most yielded the same qualitative results; manipulating the parameters in a meaningful way would have required a lot of data not directly accessible in our behavioral experiments. Furthermore, human data depend on kinematic parameters that are highly simplified in our mouse controller.

3) *Informativeness*: The usefulness and efficiency of the inhibition of return mechanism can also be ascertained by deactivating it. This is done in the model by setting  $\gamma_2$  to 0, thus canceling the effect of  $a_k^{inhib}$  on  $a_k^{reac}$ .

The consequence can be observed on the heatmaps (compare Fig. 10a and 10b). Without inhibition of return, the system converges on limit cycles of saccades from which it can hardly escape by only using the bias. Indeed, predictors prime each other through the environment and (if they belong to the same prototype) activate each other; with no inhibition of return, this mutual excitation cannot be counterbalanced. Based on the informativeness of the fixation points within the limit cycle attractor, the system may be able to reach a decision or not.

Overall, the analysis of the model dynamics leads to several predictions on human perceptual decision making, including the following: the ambiguity effect should be amplified or disappear when time constraints become more or less severe, respectively; in most cases, the perceptual boundary between two categories may not correspond to a morphing coefficient of 0.5; constraining eye movements or observed features should lead to qualitative changes in the decision dynamics and fixation patterns. Testing these predictions in further human experiments (possibly using MouseTracker in combination with a eye tracker) might help validating or refining the proposed model.

#### IV. CONCLUSIONS

Perceptual decision-making has been extensively studied in neuroscience and artificial vision research. Most experiments in neuroscience focused on simple perceptual choices, such as for instance the direction of motion of dots having different degrees of coherence [1], [2]. These studies permitted to describe the basic neural mechanisms of perceptual choice in terms of stochastic and dynamic processes [7], [3], [31]. At the same time, it is unclear if and how these mechanisms scale to

TABLE I  
PARAMETERS VALUES IN THE COMPUTATIONAL MODEL

Symbol	Value	Description
$r$	0.4	Radius of the retina (defining the visible features)
$M$	16	Number of preferred input feature orientations
$\sigma_{o_{1..M}}$	1.0	Standard deviation of the features Gaussian profile
$\sigma_{u/v}$	0.2	Standard deviation of the location Gaussian profile
$\alpha$	0.1	Decay coefficient for the inhibition of return
$\beta$	0.5	Average proprioceptive activity (normalization)
$\gamma_1$	0.1	Influence of the top-down bias on the predictors
$\gamma_2$	0.1	Influence of the inhibition of return
$a_{RSS}$	0.1	Threshold activity for the reactive/predictive control
$\lambda$	5	Inertia coefficient for the accumulation of evidences
$v_{max}$	10	Maximal speed for the mouse movement (default)

more complex perceptual tasks such as those typically studied in vision research, which usually involve complex perceptual stimuli and are addressed using feature-based representations, hierarchical and/or generative architectures [20], [22].

In this article, we propose dynamic competition as a general mechanism for perceptual decision-making of whatever complexity. The proposed model incorporates several insights of the aforementioned models of decision-making and categorization, such as dynamic competition via evidence accumulation. At the same time, our proposal has several distinguishing features. Prediction dynamics are proposed as the key principle that regulates the dynamic competition. In our model, prediction has a double role: it provides information (evidence) to be accumulated during the dynamic decision-making, and it guides the perceptual processing actively. This idea is compatible with recent theories, such as *predictive coding*, which emphasize the importance of top-down predictive processes in guiding perceptual processing [14], [16], [17], [15]; however, rather than minimizing prediction error, our system considers prediction success as a source of evidence within a dynamical system framework.

Furthermore, our model assigns a key role to *active perception* processes in the guidance of the perceptual categorizations. In this respect, it is worth noting that our active perception model is highly simplified compared to the control of human eye movements. Its main significance lies in the proposal that also evidence selection is a competitive process guided by task demands; however, evidence selection includes overt (e.g., eye movements) and covert (e.g., attention modulation) processes, which are at the moment mixed up in our model. We leave the improvement and validation of our active perception model as an open objective for future research.

Overall, our emphasis on predictions and active perception stands in contrast with the idea that evidence accumulation is a purely bottom-up process. Taken together, these elements constitute a theory of how multiple predictors compete over time for making a (perceptual) decision, and do so by both “voting” for one of the alternative and actively biasing the selection of relevant information. Consistent with embodied and sensorimotor accounts of categorization, the proposed system incorporates implicit knowledge of the categories in the form of sets of linked (feature) predictors, and contingencies

between saccades and stimuli [59], [50], [60]. By re-enacting its predictors, the system is able to successfully explore and categorize the stimuli. The model can be extended to include multi-step or probabilistic predictions that incorporate more complex relations between saccades and stimuli.

Feature-based representations are also central to our architecture, as they permit splitting the problem space and designing controllers and predictors at a manageable level of abstraction. A feature-based approach has proven useful to address perceptual tasks that are far more complex than those used in neuroscience studies [22]. By incorporating features into our model, we provide a link between dynamic competitive processes in simple perceptual tasks (e.g., using random dots) and more complex ones, suggesting that the latter consists in a competition between feature-based representations. Finally, our model incorporates response dynamics, which are rarely considered in categorization studies (for one exception, see [61]).

Although the same computational model has already been successfully applied to navigation [62] and control [63] in robotics (see also [64], [60], [65] for related proposals), it is to our knowledge the first time such distributed prediction- and competition-based system is used to model dynamic decision-making in humans. Performance of our system is assessed in a perceptual categorization of intermediate complexity (i.e., more complex than tasks usually adopted in neuroscientific studies, still less complex than the recognition of natural images). The use of a continuous measure of performance (mouse movements) permitted us to better look at the dynamics of decision during the task.

The comparison with human performance suggests that the system captures fundamental principles of decision-making, such as the sensitivity to perceptual ambiguities. Not only this affects the number of errors, but it also influences the on-line dynamics of decision, as revealed by the analysis of trajectories during task performance.

In this respect, our dynamic computational model describes decision-making as a continuous process in which perceptual processing, choice and (eye and hand) action performance all co-occur and influence one another, revealing that the embodiment of choice is indeed part and parcel of it [66]. In sum, this combined computational and human study suggests that fundamental mechanisms with which the brain implements perceptual decision-making and solves perceptual ambiguities can be extended to explain visual categorization tasks that are significantly more complex than the ones typically addressed in neuroscientific studies. We argue that these principles should guide the design of artificial systems that apply in naturalistic domains, and in turn that the advancements in the realization of artificial systems could be heuristically useful to explore the neural substrate of harder perceptual choices in living organisms.

#### ACKNOWLEDGMENT

The authors would like to thank Adam Sanborn for useful comments and for providing the experimental stimuli. The research leading to these results has received funding from

the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements 224919 (WoRHD, supporting LB), and 270108 (Goal-Leaders, supporting GP), and 231281 (EuCogII, which sponsored JCQ's visit to ISTC-CNR in Rome).

#### REFERENCES

- [1] J. Gold and M. Shadlen, "Neural computations that underlie decisions about sensory stimuli," *Trends in Cognitive Sciences*, vol. 5, pp. 10–16, 2001.
- [2] J. I. Gold and M. N. Shadlen, "The neural basis of decision making." *Annu Rev Neurosci*, vol. 30, pp. 535–574, 2007. [Online]. Available: <http://dx.doi.org/10.1146/annurev.neuro.29.051605.113038>
- [3] R. Ratcliff, "A theory of memory retrieval," *Psychological Review*, vol. 85, pp. 59–108, 1978.
- [4] M. Usher and J. L. McClelland, "On the time course of perceptual choice: The leaky, competing accumulator model." *Psychological Review*, vol. 108(3), pp. 550–592, 2001.
- [5] K.-F. Wong, A. C. Huk, M. N. Shadlen, and X.-J. Wang, "Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making." *Front Comput Neurosci*, vol. 1, p. 6, 2007. [Online]. Available: <http://dx.doi.org/10.3389/neuro.10.006.2007>
- [6] Diederich, "Dynamic stochastic models for decision making under time constraints," *J Math Psychol*, vol. 41, no. 3, pp. 260–274, Sep 1997.
- [7] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks." *Psychol Rev*, vol. 113, no. 4, pp. 700–765, Oct 2006. [Online]. Available: <http://dx.doi.org/10.1037/0033-295X.113.4.700>
- [8] M. Shadlen, R. Kiani, T. Hanks, and A. Churchland, "Neurobiology of decision making: An intentional framework," in *Better than conscious?: decision making, the human mind, and implications for institutions*, C. Engel and W. Singer, Eds. The MIT Press, 2008.
- [9] A. Tosi, G. Galati, G. L. Romani, and M. Corbetta, "Sensory-motor mechanisms in human parietal cortex underlie arbitrary visual decisions." *Nat Neurosci*, vol. 11, no. 12, pp. 1446–1453, Dec 2008. [Online]. Available: <http://dx.doi.org/10.1038/nn.2221>
- [10] D. J. Freedman and J. A. Assad, "A proposed common neural mechanism for categorization and perceptual decisions." *Nat Neurosci*, vol. 14, no. 2, pp. 143–146, Feb 2011. [Online]. Available: <http://dx.doi.org/10.1038/nn.2740>
- [11] P. Cisek and J. F. Kalaska, "Neural mechanisms for interacting with a world full of action choices." *Annu Rev Neurosci*, vol. 33, pp. 269–298, 2010. [Online]. Available: <http://dx.doi.org/10.1146/annurev.neuro.051508.135409>
- [12] R. M. Nosofsky and T. J. Palmeri, "An exemplar-based random walk model of speeded classification." *Psychol Rev*, vol. 104, no. 2, pp. 266–300, Apr 1997.
- [13] K. Lamberts, "Information-accumulation theory of speeded categorization." *Psychol Rev*, vol. 107, no. 2, pp. 227–260, Apr 2000.
- [14] M. Bar, "The proactive brain: using analogies and associations to generate predictions," *Trends in Cognitive Sciences*, vol. 11, no. 7, pp. 280–289, 2007.
- [15] G. Pezzulo, "Coordinating with the future: the anticipatory nature of representation," *Minds and Machines*, vol. 18, no. 2, pp. 179–225, 2008.
- [16] K. Friston, "A theory of cortical responses." *Philos Trans R Soc Lond B Biol Sci*, vol. 360, no. 1456, pp. 815–836, Apr 2005. [Online]. Available: <http://dx.doi.org/10.1098/rstb.2005.1622>
- [17] —, "The free-energy principle: a unified brain theory?" *Nat Rev Neurosci*, vol. 11, no. 2, pp. 127–138, Feb 2010. [Online]. Available: <http://dx.doi.org/10.1038/nrn2787>
- [18] R. P. Rao and D. H. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects." *Nat Neurosci*, vol. 2, no. 1, pp. 79–87, January 1999. [Online]. Available: <http://dx.doi.org/10.1038/4580>
- [19] Y. Bengio and Y. Lecun, *Scaling learning algorithms towards AI*. MIT Press, 2007. [Online]. Available: <http://yann.lecun.com/exdb/publis/index.html#bengio-lecun-07>
- [20] G. E. Hinton, "Learning multiple layers of representation." *Trends Cogn Sci*, vol. 11, no. 10, pp. 428–434, Oct 2007. [Online]. Available: <http://dx.doi.org/10.1016/j.tics.2007.09.004>
- [21] B. Epshtein and S. Ullman, "Semantic Hierarchies for Recognizing Objects and Parts," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007.

- [22] S. Ullman, M. Vidal-Naquet, and E. Sali, "Visual features of intermediate complexity and their use in classification." *Nat Neurosci*, vol. 5, no. 7, pp. 682–687, Jul 2002. [Online]. Available: <http://dx.doi.org/10.1038/nn870>
- [23] D. Ballard, "Animate vision," *Artificial Intelligence*, vol. 48, no. 1, pp. 1–27, 1991.
- [24] T. C. Kietzmann, S. Geuter, and P. Koenig, "Overt visual attention as a causal factor of perceptual awareness." *PLoS One*, vol. 6, no. 7, p. e22614, 2011. [Online]. Available: <http://dx.doi.org/10.1371/journal.pone.0022614>
- [25] M. Hayhoe and D. Ballard, "Eye movements in natural behavior." *Trends Cogn Sci*, vol. 9, no. 4, pp. 188–194, Apr 2005. [Online]. Available: <http://dx.doi.org/10.1016/j.tics.2005.02.009>
- [26] C. A. Rothkopf and D. H. Ballard, "Credit assignment in multiple goal embodied visuomotor behavior." *Front Psychol*, vol. 1, p. 173, 2010. [Online]. Available: <http://dx.doi.org/10.3389/fpsyg.2010.00173>
- [27] N. Sprague and D. Ballard, "Multiple-goal reinforcement learning with modular sarsa(o)," University of Rochester Computer Science, Tech. Rep., 2003.
- [28] J. K. Kruschke, "Alcove: An exemplar-based connectionist model of category learning," *Psychological Review*, vol. 99, pp. 22–44, 1992.
- [29] J. B. Freeman, R. Dale, and T. A. Farmer, "Hand in motion reveals mind in motion." *Front Psychol*, vol. 2, p. 59, 2011. [Online]. Available: <http://dx.doi.org/10.3389/fpsyg.2011.00059>
- [30] J.-H. Song and K. Nakayama, "Hidden cognitive states revealed in choice reaching tasks." *Trends Cogn Sci*, vol. 13, no. 8, pp. 360–366, Aug 2009. [Online]. Available: <http://dx.doi.org/10.1016/j.tics.2009.04.009>
- [31] M. Spivey, *The continuity of mind*. Oxford University Press, USA, 2007.
- [32] A. Resulaj, R. Kiani, D. M. Wolpert, and M. N. Shadlen, "Changes of mind in decision-making." *Nature*, vol. 461, no. 7261, pp. 263–266, August 2009. [Online]. Available: <http://dx.doi.org/10.1038/nature08275>
- [33] C. Olman and D. Kersten, "Classification objects, ideal observers and generative models," *Cognitive Science*, vol. 28, no. 2, pp. 227–239, 2004.
- [34] A. N. Sanborn, T. L. Griffiths, and R. M. Shiffrin, "Uncovering mental representations with markov chain monte carlo." *Cogn Psychol*, vol. 60, no. 2, pp. 63–106, Mar 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.cogpsych.2009.07.001>
- [35] J. B. Freeman and N. Ambady, "Mousetracker: software for studying real-time mental processing using a computer mouse-tracking method." *Behav Res Methods*, vol. 42, no. 1, pp. 226–241, Feb 2010. [Online]. Available: <http://dx.doi.org/10.3758/BRM.42.1.226>
- [36] D. Bates and M. Maechler, *lme4: Linear mixed-effects models using s4 classes*, 2009, [Computer software manual]. Available from <http://CRAN.R-project.org/package=lme4>.
- [37] R. Baayen, D. Davidson, and M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language*, vol. 59, p. 390412, 2008.
- [38] R. Dale, C. Kehoe, and M. J. Spivey, "Graded motor responses in the time course of categorizing atypical exemplars." *Mem Cognit*, vol. 35, no. 1, pp. 15–28, Jan 2007.
- [39] L. Barca and G. Pezzulo, "Unfolding visual lexical decision in time," *PLoS ONE*, vol. 7, no. 4, p. e35932, 2012.
- [40] M. J. Spivey, R. Dale, G. Knoblich, and M. Grosjean, "Do curved reaching movements emerge from competing perceptions? a reply to van der wel et al. (2009)." *J Exp Psychol Hum Percept Perform*, vol. 36, no. 1, pp. 251–254, Feb 2010. [Online]. Available: <http://dx.doi.org/10.1037/a0017170>
- [41] M. Spivey, M. Grosjean, and G. Knoblich, "Continuous attraction toward phonological competitors," *Proceedings of the National Academy of Sciences of the USA*, vol. 102, pp. 10393–10398, 2005.
- [42] R. P. R. D. van der Wel, J. R. Eder, A. D. Mitchel, M. M. Walsh, and D. A. Rosenbaum, "Trajectories emerging from discrete versus continuous processing models in phonological competitor tasks: a commentary on spivey, grosjean, and knoblich (2005)." *J Exp Psychol Hum Percept Perform*, vol. 35, no. 2, pp. 588–594, Apr 2009. [Online]. Available: <http://dx.doi.org/10.1037/0096-1523.35.2.588>
- [43] P. D. Allopenna, J. S. Magnuson, and M. K. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models," *Journal of Memory and Language*, vol. 38, pp. 419–439, 1998.
- [44] K. M. Eberhard, M. J. Spivey-Knowlton, J. C. Sedivy, and M. K. Tanenhaus, "Eye movements as a window into real-time spoken language comprehension in natural contexts." *J Psycholinguist Res*, vol. 24, no. 6, pp. 409–436, Nov 1995.
- [45] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov 1998.
- [46] M. Mermillod, P. Bonin, L. Mondillon, D. Alleysson, and N. Vermeulen, "Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation," *Neurocomputing*, no. 73, pp. 2522–2531, 2010.
- [47] M. H. Bickhard, "Function, anticipation and representation," in *Computing Anticipatory Systems. CASYS 2000 - Fourth International Conference*, D. M. Dubois, Ed. Melville, NY: American Institute of Physics, 2001, pp. 459–469.
- [48] R. Grush, "The emulation theory of representation: motor control, imagery, and perception." *Behavioral and Brain Sciences*, vol. 27, no. 3, pp. 377–96, Jun 2004.
- [49] G. Pezzulo, "Grounding procedural and declarative knowledge in sensorimotor anticipation," *Mind and Language*, vol. 26, no. 1, pp. 78–114, 2011.
- [50] J. O'Regan and A. Noe, "A sensorimotor account of vision and visual consciousness." *Behavioral and Brain Sciences*, vol. 24, no. 5, pp. 883–917, 2001.
- [51] G. L. Drescher, *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*. Cambridge, MA: MIT Press, 1991.
- [52] E. Rosch, C. Mervis, W. Gray, D. Johnson, and P. Boyes-Braem, "Basic objects in natural categories," *Cognitive Psychology*, vol. 8, p. 382439, 1976.
- [53] P. Gardenfors, *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA, USA: MIT Press, 2000.
- [54] D. L. Medin and M. M. Schaffer, "Context theory of classification learning," *Psychological Review*, vol. 85, p. 207238, 1978.
- [55] R. A. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. 2, pp. 14–23, 1986.
- [56] J. Johnson, J. Spencer, and G. Schoner, "Moving to higher ground: The dynamic field theory and the dynamics of visual cognition." *New Ideas in Psychology*, vol. 26, pp. 227–251, 2008.
- [57] J.-C. Quinton, B. Girau, and M. Lefort, "Competition in high dimensional spaces using a sparse approximation of neural fields," in *From Brains to Systems: Brain Inspired Cognitive Systems 2010*, ser. Advances in Experimental Medicine and Biology, vol. 718. Springer-New York, 2011.
- [58] M. Grosjean, J. Zwickel, and W. Prinz, "Acting while perceiving: assimilation precedes contrast." *Psychol Res*, vol. 73, no. 1, pp. 3–13, Jan 2009. [Online]. Available: <http://dx.doi.org/10.1007/s00426-008-0146-6>
- [59] L. Barsalou, "Grounded cognition," *Annual Review of Psychology*, vol. 59, pp. 617–645, 2008.
- [60] G. Pezzulo and G. Calvi, "Computational explorations of perceptual symbol system theory," *New Ideas in Psychology*, vol. 29, no. 3, pp. 275–297, 2011.
- [61] F. G. Ashby, S. W. Ell, and E. M. Waldron, "Procedural learning in perceptual categorization." *Mem Cognit*, vol. 31, no. 7, pp. 1114–1125, Oct 2003.
- [62] J.-C. Quinton and J.-C. Buisson, "Multilevel anticipative interactions for goal oriented behaviors," in *Proceedings of EpiRob'08 - International Conference on Epigenetic Robotics (Brighton, UK)*. Lund University Cognitive Studies, 2008, pp. 103–110.
- [63] J.-C. Quinton and T. Inamura, "Human-robot interaction based learning for task-independent dynamics prediction," in *Proceedings of EpiRob'07 - International Conference on Epigenetic Robotics (Piscataway, USA)*. Lund University Cognitive Studies, 2007, pp. 133–140.
- [64] G. Pezzulo and G. Calvi, "Designing modular architectures in the framework akira," *Multiagent and Grid Systems*, vol. 3, no. 1, pp. 65–86, 2007.
- [65] G. Pezzulo and D. Ognibene, "Proactive action preparation: Seeing action preparation as a continuous and proactive process," *Motor Control*, 2011.
- [66] G. Pezzulo, L. Barsalou, A. Cangelosi, M. Fischer, K. McRae, and M. Spivey, "The mechanics of embodiment: A dialogue on embodiment and computational modeling," *Frontiers in Cognition*, vol. 2, no. 5, pp. 1–21, 2011.