

# A neuroplausible computational model of vision also exhibits asymmetry in developmental category learning

Ankit Gupta (ankit0370@gmail.com)  
Department of Electrical Engineering,  
Indian Institute of Technology Kanpur, India

Devesh Kumar Singh (deveshks@iitk.ac.in)  
Department of Computer Science and Engineering,  
Indian Institute of Technology Kanpur, India

Amitabha Mukerjee (amit@cse.iitk.ac.in)  
Department of Computer Science and Engineering,  
Indian Institute of Technology Kanpur, India

## Abstract

Computational models are increasingly used to explore possible mechanisms underlying infant capability in various tasks. Often, such models do not work directly on perceptual data, but on hand-computed features of images; such models are open to the criticism that these high-level features may not be what is actually computed in the neural computation. Here we explore the feasibility of the Serre-Poggio (S-P) model which emulates the early ventral stream of the primate visual cortex, and constructs a probabilistic model of the tuned cells of the V4-IT cortex. In experiment 1, we use this system to model asymmetry in visual category learning in early infancy (e.g. cats vs dogs), and show that surprisal for the novel category is higher when habituated on CAT than on DOG. In experiment 2, we show that face habituation can be used to discriminate on full bodies. Experiment 3 demonstrates that superordinate category discriminations are easier than for the basic level. These experiments agree with earlier psychological data and partially validate the S-P model for such tasks.

## Introduction

Infant perceptual ability has been demonstrated for a wide range of visual tasks, and computational models are increasingly used to analyze possible mechanisms underlying such behaviour. However, using computational models in visual development is limited owing to the high dimensionality of visual data and the complexity of extracting meaningful structures from images. Thus, a review of computational simulation in development cites only two (out of thirty) papers for perceptual categorization (Schlesinger & Parisi, 2001).

Yet, the infant's strongest cue to abstracting from the world is perception, since its motor functions are under-developed. Between two and six months, infants demonstrate increasing ability to discriminate a number of complex concepts, primarily based on perception. Computational studies investigating infant visual perception can throw light on the internal mechanisms for such learning, and also throw light on debates such as degree of innateness, learnability, etc.

One of the areas that has attracted considerable attention in visual learning by infants is that of asymmetry in infant visual categorization (e.g. DOG vs. CAT) (Mareschal, French, & Quinn, 2000 ; Quinn, Eimas, & Rosenkrantz, 1993). This intriguing phenomenon has attracted considerable attention, and has also been addressed by a connectionist model (Mareschal et al., 2000). Here 3-6 month infants are shown a series of images from a certain category, and their preferential looking (or looking time) is analyzed for objects of the

same or different categories. Based on the novelty preference paradigm, it is expected that infants would look longer at objects from the novel category. The asymmetry result, demonstrated by Quinn and co-workers nearly two decades ago, is that infants who are habituated to cats demonstrate a preference for dogs as novel stimuli, whereas the preference is weaker when habituated on dogs and exhibited a cat. Such an asymmetry is especially surprising because it is assumed that such young infants may not have had much exposure to the classes of cats and dogs *per se*, and that any priors they form are learned only from this immediate experience.

Quinn and co-workers conjectured that the asymmetry may have arisen due to greater variation among dogs, so that some novel cats may also be accepted on the DOG schema, whereas the CAT schema is tighter and rejects most of the novel dogs (Quinn et al., 1993). This hypothesis was validated using a perceptron network by Mareschal, French and Quinn (Mareschal et al., 2000), who found that certain features had lower spread among cats than among dogs. For the simulation, they hand-computed ten traits from the images: head length, head width, eye separation, ear separation, ear length, nose length, nose width, leg length, vertical extent, horizontal extent. These features were then given as input to a three-layer perceptron, which was trained to discriminate on cats and dogs over 250 epochs. The network error rates (used internally in the backpropagation algorithm) were used as an analog for looking time, and results were shown to correlate with the original experiments on infants. Also, the gaussian distribution of these features was computed and the features for the DOG were found to have a wider variation than that for the CAT. This analysis thus corroborated the early suggestion that the asymmetry arose because the CAT class had less variability than the class DOG.

However, such an approach is open to criticism since the computation of high-level features requires that one is able to decompose the image into ear, nose, eye etc. While there is considerable evidence that the infant is sensitive to parts of the face (Perrett & Benson, 1992), nonetheless the task of mapping from a raw image to such parallel data on faces from different species remains a formidable challenge. The images, if we consider them to be the equivalent of  $300 \times 200$  pixels, have about 60,000 dimensions. The manually executed task of obtaining ten features from the image space in-

volves a mapping, say  $f_{man}$ , from the original image  $I$  to these ten features. Subsequently, the function learned by the neural network,  $f_L$ , operates a 10-dimensional feature space to obtain a binary classification into CAT,DOG. The overall process is given as

$$I \in R^D \xrightarrow{f_{man}} x_f \in R^{10} \xrightarrow{f_L} \{CAT,DOG\}$$

where  $D$  represents the dimensionality of the image, between  $10^4$  and  $10^6$  pixels, say. Clearly, learning (or knowing innately) a function such as  $f_{man}$  is not a trivial matter. Another crucial question is - which ten features to choose for the intermediate representation? Determining such a set of informative features from this heterogeneous dataset is itself injecting a considerable amount of human knowledge into the process.

A second problem with manually computing features is that only certain classes of experiments can be duplicated. In one of the experiments of the Quinn group, the infant is habituated on face images and testing is done on full bodies (Quinn & Eimas, 1996). Clearly, mapping the full-body images to the same ten feature vectors would defeat the very purpose of the experiment, and assigning some other feature class to the test images would make them incommensurate. Working directly on image data provides a mechanism for handling these aspects, and we investigate this question in our experiment 2.

### Computational model of early ventral stream

Improved understanding of function in the neuroanatomical circuits serving visual perception has led to considerable advances in computational models of the visual stream. In particular, the ventral stream from the V1, V2 and V4 areas of the visual cortex to the IT is part of the cortical computational processes thought to be responsible for object recognition or “what” questions. This pathway has attracted considerable interest from computational simulation. Tomaso Poggio’s group, working on modeling various aspects of the primate visual system, has suggested several models for this part of the pathway. The model we adopt from (Serre, Kouh, et al., 2005) seeks to replicate the gradual increase in complexity of the preferred stimuli for neurons along the ventral stream, culminating in tuned IT neurons which are believed to play a key role in object recognition (Tanaka, 1996).

The Serre-Poggio model takes a gray-scale image as input and performs four intermediate computations that combine processes in the early computation in the primate visual system. The first few hundred milliseconds in the retina and LGN involves identifying simple local maxima (e.g. center-on) processes; these are combined to obtain orientation-tuned responses. In the S-P model, these are simulated by orientation tuned simple cells (S1) which constitute the first layer, corresponding to the early part of the V1. The model has alternating layers for tuning (simple, selective) and combining (complex, invariant) computation. The S-cells are tuned for specific orientations or shapes, and the C-cells combine the responses of the neighbouring S-cells in the previous layer

using a “max” function. This gradually increases scale, specificity, and translation invariance as we move up the hierarchy. There are four layers, the first of which roughly corresponds to simple cells in V1 that are tuned to orientation (layer S1). The max for each orientation is computed over a pool of S1 cells, resulting in tolerance to position and scale within the visual field. This layer, C1 passes its output to the next tuning layer, S2 which preferentially selects differing orientations resulting in preferences for different shape primitives. It is suggested that these may correspond to cells in V4 or in the superficial layers of V2. A large number of S2 cells learn their responses based on random patches from the input. The final C2 layer used in this work maximizes across cells tuned on the same shape model, thus increasing invariance over scale and position. It is this final computed C2 output which is used in various object recognition tasks. The 4-layer hierarchy is described in Table 1.

In our experiments, we use this S-P model to compute the C2 feature vectors for different sets of habituation images. These features are used to learn probability distributions for the habituation category, which simulates the tuning of neurons in late V4 / early IT for the given visual category. The final discrimination is done by applying these tuned C2 cells to the test image  $I$ . The steps from S1 through C2 may be thought of as a mapping  $f_{vs}$ , which are then used in learning the cat, dog discrimination as follows:

$$I \in R^D \xrightarrow{f_{vs}} x_{C2} \in R^n \xrightarrow{f_L} \{CAT,DOG\}$$

Here  $f_{vs}$  is the feature mapping performed in the ventral stream. Note that unlike in the previous situation where features corresponded to conscious, declarative elements, here the features are subconscious and implicit. As in the earlier situation, the learning component of the system now has to learn a  $n$  dimensional discriminant function on this feature space, based on which it can discriminate the classes dogs and cats.

The dimension  $n$  is a function of the number of training images that random patches are sampled from; for two sets of 16 images in the training set, this results in 512 C2 units. This vector is then used for identifying a distribution. Finally, we observe that this computational model has been tuned to reflect single cell readings obtained from various sites in the visual stream of adult monkeys. Even if we assume sufficient correspondence with human visual processes, the question of relevance to the infant visual system remains. In this connection we note that there is significant evidence that infants exhibit orientation selectivity from an early stage (Wattam-Bell, 1991) and the first stage in the Serre-Poggio model recruits orientation sensitive frequency filters. What is absent in the infant are model-based priors that would be present in adults based on experience with visual stimuli of cats and dogs. However, this we do not assume, since our model is feed-forward. On this basis, Serre has described the algorithm that learns the vocabulary of tuned neurons as “developmental-like”. Indeed, our experiment can be taken as a test for the

Level	Functionalities
Simple cells (S1)	Gabor filters, 16 spatial frequencies(=scales), 4 orientations
Complex cells (C1)	Local max over a pool of S1 cells ; increase in tolerance to position and scale
Composite feature cells (S2)	Combination of V1 like complex units at different orientations
Complex composite feature cells (C2)	Local pooling over S2 units with same selectivity but slightly different positions and scales ; Same selectivity as S2 units but increased tolerance to position and size of the preferred stimulus

Table 1: Summary of layers in S-P model (based on (Serre, Wolf, & Poggio, 2005))

Gabor Filter Parameter Name	Symbol	Parameter value
Receptive field(RF) size	s	16 filters 7X7 to 37X37 (in steps of 2)
Orientation	$\theta$	$0^\circ, 45^\circ, 90^\circ, 135^\circ$
Effective Width	$\sigma$	$0.0036s^2 + 0.35s + 0.18$
Aspect Ratio	$\gamma$	0.3
Wavelength	$\lambda$	$\sigma/0.8$

Table 2: Parameters of the Gabor filter bank (Source: (Serre, Wolf, & Poggio, 2005 ; Daptardar, 2009) )

validity of such a claim for this model.

### Application to object recognition

In an object recognition task, the system is given a large number of images of different objects, and these are labelled with the relevant object categories. C2 features are computed for all the images in the training set, and these are used to train a classifier - typically, a support vector machine (SVM) (Bishop, 2007), which performs the classification. Now, given a new image, its C2 vector is computed and passed to the trained SVM which then assigns a class label to it. The model has reported 44% percentage average accuracy in discriminating more than a hundred object categories in the CALTECH-101 dataset (Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007). However, the model has some drawbacks on present (largely serial) computational architecture where the computational load is very heavy due to the dense tuning and max operation and blind feature selection.

The model replicates the shape primitives computed in the early ventral stream (upto early IT). The model relies on dense orientation data, and not on global shape - hence it would not work for silhouette experiments, say. However, many images can still be categorized, and we report the simulations for three such experiments.

**Implementation of Serre Poggio model** For this work, the S-P algorithm was re-written in C++, since the available code in MATLAB was extremely slow (Daptardar, 2009) (implementation available). OpenCV, OpenMP (libgomp), MPI (OpenMPI) and LibSVM libraries were used for this project. Table 2 describes the parameters used in the implementation of Gabor filter. The implemented model was tested on categorization tasks and performed similarly to the original work (Serre, Wolf, & Poggio, 2005).

### Modeling infant visual cognition

Visual categorization processes in infants have been studied by many experimenters including Quinn and co-workers primarily based on preferential looking paradigm (Quinn & Eimas, 1996 ; Quinn et al., 1993) though looking time or electrophysiological methods (ERPs) have also been used (Quinn, Westerlund, & Nelson, 2006). Stimuli involved images of animals and other object classes, which were often cut-out (isolated) from the background and held up on sticks, or sometimes obscured to reveal only the face (Quinn & Eimas, 1996). In many of these experiments, only one class was displayed (e.g. CAT, and the response was measured for a novel category such as DOG or some inanimate object such as CAR.

In our experiments, we focused on the preferential looking paradigms, which were modeled based on probability distributions computed on the input data set. Since there is no reason to assume otherwise, we use a gaussian model for the probabilities. The probability for novel (test) images are computed based on their C2 features  $x$ , of dimensionality  $n$ , computed from the images. It is assumed that these are sampled from a distribution  $x \sim \mathcal{N}(\mu, \Sigma)$ , estimated from the training set:

$$Pr(I_{test}|I_1, I_2, \dots, I_m) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp(-d^2) \quad (1)$$

where  $d = \sqrt{(y - \mu)^T (\Sigma)^{-1} (y - \mu)}$  and  $I_1, I_2, \dots, I_m$  and  $x_1, x_2, \dots, x_m$  are training set images and corresponding C2 feature vectors;  $I_{test}, y$  are test stimuli and its C2 feature vector;  $\Sigma$  is the covariance matrix of  $x$ . If the  $Pr(I_{test})$  is greater than some threshold  $\tau$  then the image is acceptable as a member of the habituation class, else it is considered a novel class. We report results for a threshold at which instances of the habituation class would have been accepted, and also test the effect of other thresholds.

For estimating looking time, we assume that looking time is proportional to surprisal, - an information-theoretic notion and corresponding to the amount of information present in a single observation (Bishop, 2007). Surprisal is often correlated to looking-time in cognitive models, e.g. in psycholinguistics (Levy, 2008).

Thus, an observation that is close to expectation generates low surprisal, and highly deviant (novel) data generates higher surprisal. Surprisal for novel test stimuli  $I_{test}$  with C2 feature vector  $y$  is defined as

$$surprisal(I_{test}) = -\log \{Pr(I_{test}|I_1, I_2, \dots, I_m)\} \quad (2)$$

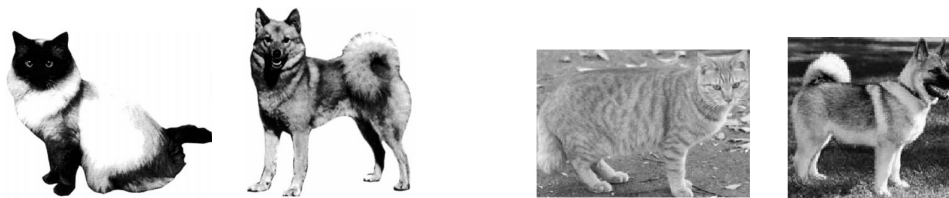


Figure 1: Sample dataset: first 2 images from Quinn et al. (2001) and last two were from *CAT-DOG database* (2011)

Hence, we have  $surprisal(I_{test}) = d^2 + c$ , where  $c$  is a constant estimated from the  $\Sigma$ .

### The experiments

We conducted three experiments on the lines of three of the experiments reported by Quinn and his group. In the first experiment, the system was trained on 16 full-body images from the categories CAT or DOG, and it computes a separate gaussian model for each of these classes on the C2 features. Next, it was shown a pair of individual images one of CAT and one of DOG. Unlike with human infants, while exhibiting images, the learning is switched off, so that each episode effectively replicates one experiment with an infant (albeit, the same one). Modeling looking time based on surprisal, our results are compared with experiments done on real infants with full-body images of these animals from (Mareschal et al., 2000). We also report the incremental error rates in assigning test images to the habituation (training) category or novel category.

Experiment 2 replicates the face to whole body learning referred to above (Quinn & Eimas, 1996). Here, we train on 16 face images from these two classes, and test on full-body images of CAT and DOG.

Experiment 3 considers hierarchical categorization (Quinn, 2004 ; Behl-Chadha, 1996 ; Quinn et al., 2006). Here we seek to demonstrate that distinctions between superordinate categories (e.g. CAR vs CAT) is much more robust than basic categories such as CAT vs DOG.

### Stimuli

We chose CAT, DOG image exemplars similar to those used by (Quinn & Eimas, 1996 ; Quinn et al., 1993)), except that in the original experiments, the figure objects were cut out from the background so as to avoid any distractors. This assumes segmentation capability that an infant may or may not have. We did some early tests which showed the system doing almost as well on un-cropped images (without too much background) as with cropped; so all results are reported on un-cropped images.

Images of the animal standing sideways, on all fours, were selected for the full-body experiment, while for face only case, images with the animal facing the viewer were chosen. The image sets are available online (*CAT-DOG database*, 2011).

The other issues are that a) a large number of image instances are needed for category learning, and b) selection of

exemplars must be such that any individual instance does not affect the overall performance of the model. Figure 1 shows the sample database used in earlier experiment and in our experiment.

### Experiment 1: Asymmetry in category learning

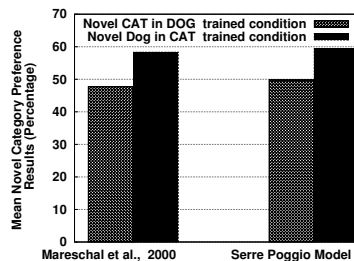


Figure 2: Comparison of mean percentage performance of the asymmetry experiment done on infants (Mareschal et al., 2000) with S-P model

In this experiment, a set of habituation images  $I_{training}$  was used to learn a distribution for a single category (in the C2 feature vector space). Now, given two test images  $I_1$  and  $I_2$ , the image with higher surprisal will correspond to lower conditional probability  $Pr(I_i | I_{training})$  and which means high looking time. Different sets of 16 images from a single category are randomly selected as  $I_{training}$  for each trial of familiarization phase, resulting in an estimated distribution  $\mathcal{N}(\mu, \Sigma)$ . For testing, we choose all possible pairs of new test images objects from a set of 8 CAT and 8 DOG images. After training with the CAT category, situations where DOG images are preferred are compared with the preference for the CAT images. The result (average over 5 runs), compares well with the original infant experiment (Mareschal et al., 2000) (Figure 2).

Further, we tested the discrimination between these categories based on different training sets with 16 images of both categories, tested with a mix of cat and dog images. Unlike the conditional probability maximization method in the earlier single-category approaches, we use an SVM now to discriminate among the test images. When the model was trained on 16 CAT images against 16 images from a mix of animals, its average categorization accuracy on 8+8 test images was 65% for CAT and 55% of DOG images were found to be novel (Figure 3). Conversely, when trained on DOG, the

test data categorization was 75% for dogs, and only 20% of cats were found to be novel - i.e. 80% of the cats were accepted as dogs. To determine the stability of this result, we tested the result by choosing different subsets for training vs test images; the results of these tests, summarized in Fig. 4 show a consistently better performance in recognizing CATs. This demonstrates that a) the model is able to discriminate the novel full bodies after learning from few training exemplars, and b) there is a high degree of asymmetry in performance, as in the original experiment. This suggests that an SVM classifier in some respects may be behaving similar to mechanisms in infants, though this is far from saying that infants are actually using such mechanisms.

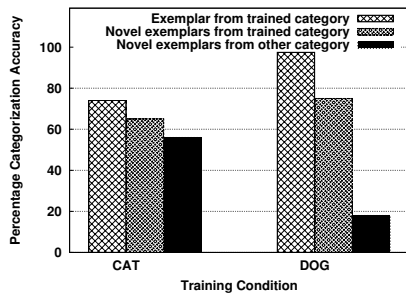


Figure 3: Categorization performance of S-P model for full body images of CAT and DOG.

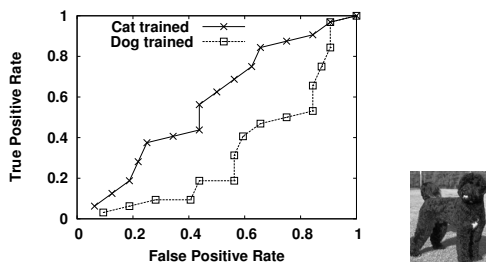


Figure 4: (Left) True positive rate vs False positive rate, across all discriminating thresholds. The top left corner (0,1) implies perfect discrimination, and cat results are consistently closer to it, hence better. (Right) A DOG image that has very high acceptability in the CAT category.

The observed asymmetry in categorization performance for the novel category implies that the model identifies DOG more accurately after training on CAT, as compared to CAT after training on DOG. We also observe that high variability in the dog category is indicated by some individuals who tend to have greater acceptability in the cat category (lower conditional probability) than most cats (Fig. 4.b). On the whole, this similarity to infant categorization results suggests that the model is able to capture the relevant attributes of infant visual learning.

In order to consider the effect of the threshold, we compute the discrimination over all possible thresholds, from the

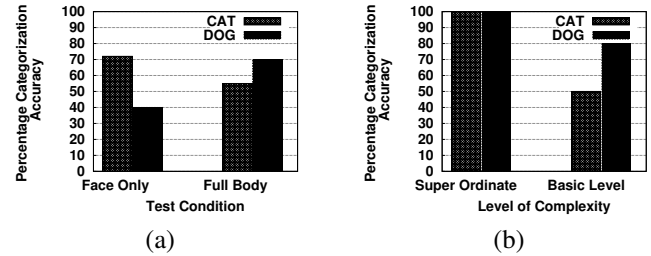


Figure 5: Experiments 2 and 3: a. Role of facial information in categorization of full body images of CAT and DOG; b. Better categorization performance at superordinate (CAR vs ANIMAL) than at basic level (CAT vs DOG).

lowest discriminating value to the highest. We report the True Positive Rate (percentage of novel detections that are actually novel) vs the False Positive Rate (percentage of the non-novel detections that were actually novel) for different thresholds (Fig. 4.a). Regions for very high and very low thresholds are noisy, but in the middle part, curves closer to the top left corner have better discrimination. Thus, the data trained on CAT is stronger for a range of feasible thresholds.

### Experiment 2: Face vs. Full-body

Quinn et al. (1993) ; Quinn et Eimas (1996) ; Quinn et al. (2001) showed that shape or facial/head informations of cat and dog are sufficient for 3-4 month old infants to form categorical representations that can discriminate based on full body images, but also demonstrates similar asymmetry. Experiment 2 tests this, by training the model on face exemplars and testing on novel face images. When habituated on CAT, 72.5% of novel category data (dogs) are preferred, whereas for DOG, only 40% for novel images are preferred (Fig. 5.a). In the next experiment, the training was on the face exemplars but the test image exemplars was novel full body exemplars. In this case the model was able to categorize 70% of the novel dogs and 55% of the novel cats accurately. The high categorization accuracy in both experiments support the claim that the head region provides significant information for learning individuated category representations for cats and dog. Further ramifications of this result, in a system without any prior experiences for faces of any kind, are suggested in the conclusion.

### Experiment 3: Hierarchical categories

When categories are organized in hierarchies, behavioral and electrophysiological studies reveal that superordinate categories are easier to learn (Quinn, 2004 ; Behl-Chadha, 1996 ; Quinn et al., 2006). We tested this result on images at the superordinate level (CAR vs ANIMAL) and at basic level (CAT vs DOG). While the model was able to discriminate completely at the upper level, at the basic level, only 50% of CAT and 80% of DOG images are identified. (Fig. 5.b). This also tallies with the original results, thus suggesting that this computational mechanism may be behaving similarly to some aspects

of infant category learning.

## Conclusion

In this work we have shown that a computational model, originally constructed based on the feedforward behaviour of the primate visual cortex, is able to replicate the infant visual response in several scenarios, working directly on image data as opposed to hand-computed features. This procedure enables a number of avenues for further testing of many other aspects of infant (and adult) learning from image data. It also lends further weight to the early suggestion that this asymmetry is due to a greater variation in the dogs than in cats.

The primary question we sought to answer is if the S-P computational model exhibits behaviour similar to infant visual cognition, given that it was initially modeled on adult primates, albeit in a feedforward manner. The results of our experiments appear to lend some support for this position, qualified by the absence of maturational and other aspects.

Another aspect on which the results may have some slight bearing is the debate on whether the capability for face recognition is innate. Though there is broad agreement that infants have some degree of face preference at birth, whether this is genetically encoded or not appears to be a matter of some debate (Johnson, 2001). The results of experiment 2 demonstrate the effectiveness of the face-only test in a system which has absolutely no priors for faces. This suggests that faces may be recognized early as information rich visual elements, and hence attended to early on. This would result in face information being quickly assimilated into cortical structures. Thus, the experiment appears to indicate some degree of learnability for face competence.

More than the immediate relevance to these experiments on CAT-DOG asymmetry, we feel that the technique presented here - the first end-to-end computational model to simulate the visual behaviour of infants - may have broader implications. The capability for, and mechanisms of object recognition have a fundamental role in behaviour, and constructing better models for it have not only improved our understanding of cognition but also empowered a rich line of investigation in computational vision. Unlike earlier attempts at simulation where the details of the visual processing had to be approximated, now the internal mechanisms as posited by this model can be assessed for its relevance to a wide range of visual developmental phenomena. Thus although this computational simulation holds considerable interest *per se*, the future value of this work may lie in the possibility of increasingly realistic simulation of developmental visual phenomena.

## Acknowledgments

We are thankful for help received from Sourabh Daptardar regarding running the C++ version of the S-P algorithm coded by him.

## References

Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*,

60(2), 105 - 141.

Bishop, C. M. (2007). *Pattern recognition and machine learning* (1 éd.). Springer.

Cat-dog database. (2011). (IITK Vision Group : <http://www.cse.iitk.ac.in/~vision/>)

Daptardar, S. (2009). *Explorations on a neurologically plausible model of image object classification*. Masters dissertation, Dept CSE, IIT Kanpur. (code: <http://code.google.com/p/object-classification-experiments/>)

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A Connectionist Account of Asymmetric Category Learning in Early Infancy. *Developmental Psychology*, 36(5), 635-645.

Perrett, M. W., D. and Hietanen J.K. and Oram, & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. London Ser.*, 335(1273)(1), 23-30.

Quinn, P. C. (2004). Development of subordinate-level categorization in 3- to 7-month-old infants. *Child Development*, 75, 886-899.

Quinn, P. C., & Eimas, P. D. (1996). Perceptual Cues That Permit Categorical Differentiation of Animal Species by Infants. *Child Psychology*, 6(1), 189 - 211.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, 22(4), 463-475.

Quinn, P. C., Eimas, P. D., & Tarr, M. J. (2001). Perceptual Categorization of Cat and Dog Silhouettes by 3- to 4-Month-Old Infants. *Child Psychology*, 79(1), 78 - 94.

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural Markers of Categorization in 6-Month-Old Infants. *Psychological Science*, 17(1), 59-66.

Schlesinger, M., & Parisi, D. (2001). The Agent-Based Approach: A New Direction for Computational Models of Development. *Developmental Review*, 21(1), 121-146.

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., Poggio, T., et al. (2005). *A theory of object recognition: Computations and circuits in the feedforward path of the ventral stream in primate visual cortex* (Rapport technique N° CBCL-259). MIT.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Tr. on Pattern Ana. and Mac. Int.*, 29, 411-426.

Serre, T., Wolf, L., & Poggio, T. (2005). Object recognition with features inspired by visual cortex. *CVPR'05 - Volume 02*, 994-1000.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual review of neuroscience*, 19(1), 109-139.

Wattam-Bell, J. (1991). Development of motion-specific cortical responses in infancy. *Vision Research*, 31(2), 287 - 297.