MODELLING ASYMMETRIC INFANT CATEGORIZATION WITH THE REPRESENTATIONAL ACUITY HYPOTHESIS

G. WESTERMANN* AND D. MARESCHAL
Centre for Brain and Cognitive Development
School of Psychology, Birkbeck College
Malet Street
London WC1E 7HX, UK

We present a neural network model that accounts for an observed asymmetry in the categorization of cats and dogs in 3–4 month old infants. The model establishes a link between infant behaviour and mechanisms of cortical processing. Based on developmental change in the cortex the model predicts behavioural change in infants between 3 and 10 months of age.

1. Introduction

What are the brain mechanisms underlying categorization? Computational models of categorization in infancy can shed light on this question in multiple ways. They have not only been able to account for different aspects of categorization behaviour [2–5] but they have also been successful in demonstrating how behavioural change can occur as the outcome of developmental changes in the cortex [11]. These models generally simulate infant behaviour in preferential looking tasks. In these tasks, infants are shown a sequence of pictures (or real objects) from one category (for example, pictures of cats) and the time the infants spend looking at these stimuli is measured. With continued exposure to these familiarization stimuli looking time normally decreases and the infants habituate to them. In a subsequent test phase, infants are presented with two pictures, one from the old category (e.g., another cat), and one from a new category (e.g., a dog). Infants prefer to look at novel things, and if they look, for example, longer at the novel dog but not the novel cat, it is assumed that they have formed a category of cats that includes the novel cat, but excludes the dog.

*current address: Department of Psychology, Oxford Brookes University, Oxford OX3 0BP, UK. email: gwestermann@brookes.ac.uk
Infant behaviour in a variety of such categorization tasks has been successfully modelled in auto-encoder neural networks [2–5]. These are simple feed-forward backpropagation networks that learn to reproduce the input on the output side. Normally these models have a hidden layer that is smaller than the input and output layers, and the idea is that this bottleneck forces the model to develop a structured internal representation of the presented stimuli. The models are trained, like infants, on a set of familiarization stimuli and subsequently presented with test items. The network error (i.e., discrepancy between the stimulus representation and the network output) is seen as analogous to infant looking time (Fig. 1). This analogy is based on the idea that infant looking time is determined by the time necessary to build an accurate internal representation of an object by cycling through an encoding/comparison/adjustment cycle, and the more novel an object is, the longer this time will be [9]. In auto-encoder models, a more novel object will generate a higher network error which likewise would require more adaptation cycles than a familiar object.

In the categorization of cats and dogs, a striking asymmetry has been found [8]: when 3–4 month old infants were familiarized on pictures of cats, they subsequently preferred a novel dog over a novel cat, indicating that they had formed a category of cats that included the novel cat and excluded the novel dog. However, when infants were familiarized on dogs, they showed no preference for either a novel cat or a novel dog, indicating that their “dog” category included novel cats as well.

This and related results have been modelled in a series of simulations
with auto-encoder neural networks [2; 3; 5]. In these simulations, the cats and dogs were described in terms of ten key features (leg length, ear separation, ear length, nose length, nose width, head length, head width, eye separation, body length, body height). The models were trained on either 12 randomly selected cats or dogs, and were then tested on a novel cat and a novel dog. Mareschal et al. [5] found that their model displayed the same asymmetry as the 3–4 month old infants: after being trained on cats, a novel dog produced higher error than a novel cat. By contrast, after being trained on dogs, there was no significant difference between the errors for novel cats and dogs. Based on the model, Mareschal et al. [5] argued that the behaviour of the infants was determined by the perceptual characteristics of the stimuli: the distribution of the stimulus features showed less variance for cats than for dogs, and, importantly, the cat feature values were subsumed under the dog feature values. Mareschal et al. concluded that early infant categorization could be explained by perceptual bottom-up processes and was independent of higher-level category knowledge. Subsequent simulations and infant experiments have provided further support for this claim [2; 3; 7].

In this paper we present simulations of these experiments with a new model, the Representational Acuity Hypothesis Model [11]. These simulations present several improvements over the conventional auto-encoder model. In order to avoid catastrophic forgetting, the auto-encoder model had to be trained in batch mode, that is, all familiarization items were presented to the model before a weight update was made. However, this procedure does not correspond to the infant familiarization where stimuli were presented individually and it can be assumed that infants adapted to each presented stimulus when it was presented. Second, the auto-encoder modelled a static snapshot of the developmental stage of infants without accounting for transitions between behaviours. In fact the same model was argued to account for 10-month olds’ behaviour in one study [4] and for the behaviour of 3–4 month olds in another [5]. Finally, in our model we want to make a more explicit link with cortical processing and show how developmental change can be explained by changes in cortical processing.

The Representational Acuity Model is based on the same principles as the previous auto-encoder, encoding familiarization stimuli and using the model error as an analogy to infant looking time in behavioural experiments. However, the internal structure of the model is different, leading to its ability to account for mechanisms of developmental change and to the identification of cortical mechanisms that underlie this change.
2. The Representational Acuity Hypothesis

The Representational Acuity Hypothesis (RAH) [11] was developed to account for the developmental change in categorization behaviour of infants from 4 to 10 months of age. This change has been argued to involve a shift from processing individual features at 4 months of age to processing relations between features at 10 months of age [12]. In these experiments, infants were familiarized on a set of animal pictures that varied along three feature dimensions (foot type, body type, tail type). Occurrence of individual features was correlated, that is, for example, a certain tail would always co-occur with a certain body. Infants were then tested on three animal pictures: one observed the feature correlation of the familiarization set, one had completely novel features, and one had previously used features, albeit in novel combinations that violated the correlational structure of the familiarization stimuli. In these experiments, 4-month olds did not look preferentially at this “switched” animal, whereas 10-month-olds did. This result was seen as evidence that 10-month olds had processed the correlations between features in the familiarization data and therefore considered the “switched” animal as novel, whereas 4-month olds had not [12].

We explained these results in our model based on the RAH [11]. We argued that objects are represented in higher cortical areas in terms of salient features. A neuron on this cortical “representational map” responds when the featural description of an object falls within its receptive field. In analogy with the development of visual acuity in the first months of life we argued that representational acuity develops based on a decrease of neuronal receptive field size during development. In young infants receptive fields are large and neurons respond to a variety of featural descriptions of objects. During development these receptive fields shrink and the neurons become tuned to more specific objects. Whereas rapid adaptations of receptive fields in response to particular stimuli have been reported (e.g., [10]), the RAH is concerned with the more gradual decrease in receptive field size that occurs over a timescale of months during early development (e.g., [1]).

We modelled the RAH with an auto-encoder neural network that had a hidden layer of units with Gaussian activation functions (in effect, an auto-encoder radial basis function network). The hidden layer modelled the representational cortical map. The stimuli used in the infant experiments [12] were encoded in terms of their features and presented to the model which was trained to reproduce these inputs on the output side. To
model the development of representational acuity, the model was trained with different widths of the hidden Gaussian activation functions. Younger infants were modelled with large receptive fields, and older infants with small receptive fields. Like in the conventional auto-encoder [4], network error for test items was considered analogous to infant looking time for these test items.

Training the model with different receptive field sizes accounted for the behavioural shift that was observed with infants of different ages [11]. With large receptive fields, the model displayed the same error for the previously seen and the “switched” test stimulus, corresponding to the behaviour of 4-month olds, whereas with small receptive fields, error for the switched item increased to be as high as that for the novel test item, corresponding to the behaviour of 10-month olds.

3. Modelling Asymmetry in Categorization

In this paper we extend our simulations with the RAH model as a biologically more plausible alternative to the conventional auto-encoder model and show how the categorization behaviour of 3–4 month old infants can be explained with recourse to the development of cortical receptive fields. Further, by varying the size of receptive fields in the model we can make prediction about the categorization behaviour of older infants.

3.1. Cat–Dog Asymmetry

The first study addressed the original results of cat–dog asymmetry [8] that have previously been modelled using the auto-encoder model [5]. We used the same data as this earlier model. The RAH auto-encoder had 10 linear input and 10 linear output units. 400 Gaussian hidden units were randomly placed in the 10-dimensional input space. To simulate young infants with large cortical receptive fields a receptive field size of 3.8 (1.2√d with d = 10 the dimensionality of the input space) was used. For cat-familiarization trials, the model was trained by presenting 12 randomly selected cats in pairs of two for 1000 iterations (weight updates) before proceeding to the next pair. For dog-familiarization trials, the model was likewise trained on 12 randomly selected dogs in pairs of two. Weights from the hidden to the output units were adapted at each iteration with the perceptron learning rule. Receptive field positions (that is, input-hidden weights) were not adapted. After training, the model was tested on a novel cat and a novel dog. Each simulation was run 50 times with different random weights and
hidden unit positions.

Fig. 2A shows the result of this simulation. When the model was trained on cats, test error for a novel dog was significantly higher than for a novel cat. By contrast, when trained on dogs, test error for novel cats and novel dogs was not significantly different. This result models the experimental data from 3–4 month old infants [8].

Figure 2. Simulation of the asymmetry between cat and dog categorization with large (A) and small (B) receptive fields. Standard error bars are also plotted. Results are averaged over 50 simulations each.

Why does the model display the asymmetry in categorization? As shown by Mareschal et al. [5], in most of the relevant feature dimensions the features for cats show less variability than those for dogs and are subsumed under the dog features. In effect, all cats look rather similar to each other, whereas dogs can look quite different from each other. On the representational feature map this means that the neurons that are activated by representations of cats lie within the area of neurons that are activated by dogs (Fig. 3A). As a consequence, when the model is trained on dogs the cat receptive fields are also (partially) tuned to the presented stimuli because the large receptive fields for cats and dogs overlap in this region (Fig. 3B). Subsequent presentation of a cat therefore does not lead to a higher error than presentation of a dog. However, when the model is trained on cats, only a small subset of the dog receptive fields are tuned (Fig. 3C), and subsequent presentation of a novel dog leads on average to higher error than a novel cat.

The asymmetry in the model exists because the receptive fields of neurons responding to cats and dogs overlap. This overlap is due to their large size, simulating an early stage of infant development. We have argued that
Figure 3. A. An illustrative example of the distribution of cats and dogs on the representational map. B. Cat receptive fields that are tuned by responding to dogs. C. Dog receptive fields that are tuned by responding to cats. For illustrative purposes here the representational map is assumed to represent only two feature dimensions (instead of ten).

the decrease in receptive field size in the cortex is a mechanism that leads to behavioural change in infant categorization [11]. In order to model the categorization of cats and dogs by older (10-month old) infants, we repeated the simulation exactly as described above, with the only difference that the size of the hidden unit receptive fields was reduced to 0.63 \(0.2\sqrt{d}\). The result of this simulation is shown in Fig. 2B. The categorization asymmetry has disappeared: when trained on cats, as before, error is higher for a novel dog than for a novel cat. However, now there is also a higher error for novel cats when the model has been trained on dogs. The model thus predicts that the category asymmetry in 3–4 month old infants disappears with increasing age, due to the decrease in receptive field size on a representational cortical map. This result has not been explicitly tested in preferential looking experiments, however, object examination studies have found that infants form exclusive cat and dog categories by 11 months [6].

3.2. Reversing the Asymmetry

To verify the claim that early infant categorization is based on perceptual bottom-up processes and depends on the feature distribution in familiarization and test stimuli, French et al. [3] showed that the cat-dog asymmetry could be reversed by carefully selecting and slightly distorting pictures of cats and dogs so that now, cat features had a broader variance than dog features and subsumed the dog features. They predicted that with this new stimulus set, infants should show a reversed asymmetry, preferring a novel cat when familiarized on dogs, and showing no preference when familiarized on cats. Their prediction was confirmed in experiments with 3–4 month olds as well as in simulations with the auto-encoder model.
The RAH model was also able to account for these results. Fig. 4A shows the network error when trained on the new data in the same way as above, with large receptive fields. Now, when trained on dogs, network error is higher for a novel cat than for a novel dog, but when trained on cats, there is no significant difference between the errors.

Figure 4. Modelling the reversal of asymmetry with large (A) and small (B) receptive fields. Standard error bars are also plotted. Results are averaged over 50 simulations each.

Modelling older infants in the asymmetry reversal condition by using smaller hidden unit receptive fields (Fig. 4B) yielded the surprising result that here the asymmetry did not disappear. The reason for this result is that familiarization on cats did not decrease the error for novel cats in testing. This might be due to the fact that the cats in the asymmetry-reversal condition are more varied than the dogs in the asymmetry condition. Infant studies with 10-month olds will have to test this prediction.

3.3. Eliminating the Asymmetry

In a final study \[2\], French et al. eliminated the asymmetry in young infants by again choosing carefully and distorting cat and dog pictures in a way that their feature distributions did no longer overlap. Infant experiments and simulations with the auto-encoder model showed that with this data set, the asymmetry did indeed disappear, providing further strong evidence for the bottom-up feature-based view of infant categorization.

We modelled this result in the RAH model by training it as before, but this time with the non-overlapping stimuli. As before, to model younger infants we used a large receptive field size. The results of this simulation are shown in Fig. 5. In this condition, even with large receptive fields, the
model shows a symmetric categorization, with novel dogs preferred over (resulting in higher error than) novel cats when trained on cats, and novel cats preferred over novel dogs when trained on dogs.

![Graph showing network error (looking time) for trained on cats and dogs](image)

Figure 5. Modelling the suppression of asymmetry in younger infants with large receptive fields. Standard error bars are also plotted. Results are averaged over 50 simulations.

4. Discussion

The simulations reported in this paper provide further support for the Representational Acuity Hypothesis [11]. The RAH was implemented in a simple neural network model that, in varying the size of the hidden unit receptive fields, makes direct reference to developmental changes in the brain to produce behavioural change in infant categorization. The model can account for observed behaviours in the categorization of cats and dogs in 3–4 month old infants, and it makes predictions about the behaviour of 10-month olds. Together with the simulations of behavioural change leading from the processing of individual features to correlations between features [11], the RAH provides a parsimonious account of the link between brain and cognitive development in early infant categorization.

Acknowledgments

This work was supported by European Commission RTN grant HPRN-CT-2000-00065 and by ESRC (UK) grant R000239112.

References

1. B. O. Braastad and P. Heggelund. Development of spatial receptive field organization and orientation selectivity in kitten striate cortex. *Journal*
REFERENCES

1. M. Ramachandran, V. Albert, and R. M. French. The role of bottom-up processing in perceptual categorization by 3- to 4-month old infants: Simulations and data. (under review).


