The Impact of Labels on Visual Categorisation: a Neural Network Model

Valentina Gliozzi

(gliozzi@di.unito.it)
Dipartimento di Informatica, Universitá di Torino,
Corso Svizzera 185, 10149 Torino, Italy

Julien Mayor and **Jon-Fan Hu** and **Kim Plunkett**

(julien.mayor, jon-fan.hu, kim.plunkett @psy.ox.ac.uk)
Department of Experimental Psychology, University of Oxford
South Parks Road, Oxford OX1 3UD, UK

Abstract

We propose a computational model of the impact of labels on visual categorisation. The proposed model is based on self-organising maps. The model successfully reproduces the experiments demonstrating the impact of labelling on infant categorisation reported in Plunkett, Hu, and Cohen (2008). Two architectures are explored. Both mimic infant behaviour in both the familiarisation and testing phases of the procedure, using a training regime which involves only single presentations of each stimulus. The model reproduced these results in the absence of a explicit teaching signal and predicts that the observed behaviour in infants is due to a transient form of learning that might lead to the emergence of hierarchically organised categorical structure.

Keywords: self-organising maps, connectionist modeling, categorisation, lexical development.

Introduction

We present a computational investigation of the role of language on visual categorisation. Our model reproduces recent findings by Plunkett et al. (2008) that show that labels have an impact on the way in which visual stimuli are categorised. In a series of five experiments, Plunkett et al. (2008) demonstrate that the manner in which 10-month-old infants perceive the visual similarity of objects depends upon whether infants are familiarised with the objects in the presence or absence of labels, and upon the labelling contingencies during familiarisation, i.e., how the labels correlate with category membership.

The model provides a mechanistic account of this dramatic effect of labels on infant visual categorisation. It is based on self-organising maps, proposed by Kohonen (1997) as a biologically plausible alternative to error-driven algorithms, such as back-propagation. In the 'familiarisation phase' of their experiments, Plunkett et al. (2008) presented infants with a set of eight line drawings, accompanied by novel words in some conditions. Similarly, in our simulation, neural networks are familiarised with a set of eight inputs that encode the stimuli presented to the infants. In contrast to previously proposed connectionist models, and similarly to infants, each of the stimuli is presented to the network only once. After the familiarisation phase is complete, infants are tested with new drawings, presented without labels. Following the well-established novelty preference procedure, infant categorisation of the stimuli in the familiarisation phase is assessed

by measuring their looking time at the different test drawings. Similarly, network categorisation is assessed by measuring its 'quantisation error' when presented with new test stimuli that only contain a visual component. If we consider the error as the analogue of the looking time in infants, following Mareschal and French (2000) and Westermann and Marechal (2004), we find that our model reproduces infant behaviour during testing. Furthermore, there is a correspondence between infant behaviour and the model's during the familiarisation phase: to a large extent infant looking time and network error rate during familiarisation follow the same trend. We argue that this match contributes to the psychological plausibility of our model.

The model comes in two variants: one variant is very simple and only contains a single self-organising map, whereas the second variant is more complex and contains three self-organising maps, namely a visual, an acoustic and a global map. The two variants replicate all five experiments reported in Plunkett et al. (2008).

The Role of Labels on Infant Categorisation

According to Plunkett et al. (2008), labels have a strong impact on how infants form categories. In two experiments, they replicate some results of Younger (1985) on visual categorisation, in which labels are not presented. During a familiarisation phase, 10-months old infants are exposed to a set of 8 visual stimuli. These stimuli are cartoon drawings of animal-like objects defined along four dimensions, corresponding to the length of the neck and legs, to the ears' orientation and the size of the tail. Under the so-called 'Broad Condition' (used in Exp. 1) infants are exposed to cartoons whose feature-values freely vary, and which can be organised in a single large cluster centered on a prototype. Under the so-called 'Narrow Condition' (used in Exp. 2), infants are shown cartoons which are variants of two prototypes and that can be grouped in two distinct clusters. In other words, under the Broad Condition, the drawings invite infants to form a single category, whereas under the Narrow Condition they invite them to form two distinct categories. Images from the two conditions are shown in Figure 1.

Once the familiarisation phase is complete, category formation is assessed by measuring the looking time of infants when tested with two different kinds of cartoon drawings: the
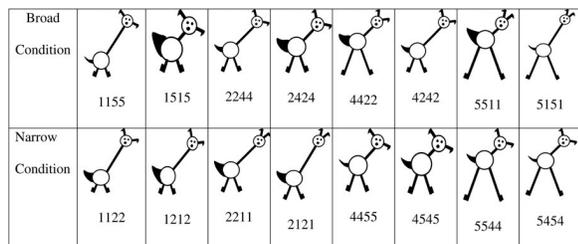
Figure 1: Cartoons shown to infants during familiarisation.

'average' drawing (the central tendency of all the drawings presented during familiarisation) and a 'modal' drawing (at the edges of the drawings presented during familiarisation, and close to one of the two prototypes used in the Narrow Condition). Following the well-established novelty preference procedure, looking time is taken to be a measure of infant surprise at the new stimuli. If infants look longer at a modal cartoon than at the average one, this is taken as a sign that they have formed a single category for all cartoons; in contrast, if they look longer at the average stimuli than at the modal ones, this is taken as a sign that they have formed two distinct categories. Results show that when familiarised with the Broad Condition, infants form a single category containing all the stimuli, whereas when familiarised with the Narrow Condition, they form two distinct categories.

In order to assess the impact of labels on categorisation, Plunkett et al. (2008) consider whether labels affect the categorisation of the visual stimuli, as described in the first two experiments. To this end, three more experiments are conducted. In all three experiments, infants are exposed to the same visual stimuli as those used in Exp. 2, namely the visual stimuli from the Narrow Condition. In contrast with Exp. 2, the visual stimuli are accompanied by acoustic labels. In Exp. 3, two labels are presented, one for each visual category; in Exp. 4 two labels are presented, randomly associated with the visual stimuli; in Exp. 5 a single label is presented with all the images. As for Exps. 1 & 2, category formation is assessed by measuring the looking time of infants when tested with modal and average stimuli in silence.

As with Exp. 2, in Exp. 3 infants formed two categories, demonstrating that labels do not interfere with visual categorisation, when they are correlated with the visual categories. In Exp. 4 however, in which labels are not correlated with the visual categories, no category is formed, as indexed by a lack of preference for either the modal or average stimuli. This experiment indicates that a decorrelation between labels and visual inputs has a disruptive effect on the process of category formation. Finally, in Exp. 5 one single category is formed, as indexed by a preference for the modal stimuli during testing — an analogous result to Exp. 1. This result emphasises the importance of labels in category formation: in this case, a single label alters the manner in which visual stimuli are categorised. Indeed, the same visual stimuli without labels lead infants to form two categories (as in Exp. 2),

whereas when accompanied with a single label they give rise to a single category. Each experiment was carried out with 24 infants. The average ratio between looking time at the average stimulus and looking time at one of the modal stimuli throughout the experiments is reported in Table 1[1].

Exp.	LT average:modal	ratio LT average/total (SD)
1	2.67s:3.35s	44.2 % (9.6)
2	3.43s:2.72s	56.1 % (11.8)
3	2.30s:1.90s	55.1 % (9.8)
4	2.14s:2.04s	50.3 % (12.2)
5	1.88s:2.14s	45.8 % (7.7)

Table 1: Infant looking time during testing.

Related Work

To the best of our knowledge, there are no previous computational models of the dramatic effect of labels on visual categorisation described by Plunkett et al. (2008). Plunkett, Sinha, Møller, and Strandsby (1992), and Mirolli and Parisi (2005) hypothesize that language might accelerate the clustering of visual images and improve the quality of visual categorisation. This hypothesis is much weaker than the hypothesis we are considering here. Furthermore, these earlier models are based on traditional backpropagation networks, which are widely recognised as biologically implausible (e.g. Crick, 1989). The same drawback applies to Mareschal and French (2000) and Westermann and Marechal (2004) who described successful simulations of the Younger (1985) experiments reported by Plunkett et al. (2008).

More biologically inspired models have been proposed to simulate aspects of infant linguistic development different from those we are dealing with here. For instance, Li, Zhao, and Whinney (2007) propose a model to account for various phenomena connected to early vocabulary acquisition, such as the vocabulary spurt, frequency and length effects; Mayor and Plunkett (2008) model the generalisation of word-object associations from a single labelling event (taxonomic constraint); Miikkulainen (1997) accounts for category-specific impairments in language.

Our Model

Encoding of the Stimuli

For the purposes of our simulations, we have encoded the visual and acoustic stimuli by means of value vectors. For both kinds of stimuli, we have used input vectors with four dimensions (e.g. $[0.27, 1, 0.8, 0.33]$). Following Mareschal and French (2000), in the case of visual stimuli each value in the input vector corresponds to one feature in the cartoons presented to infants. In the Narrow Condition, the visual map receives input vectors that can naturally be divided into two categories. This parallels infants natural clustering into two categories of the images in the Narrow Condition, as described in Plunkett et al. (2008). In contrast, in the Broad Condition, the

---

[1]Testing time in Exps. 1 & 2 is 10 sec. In Exps. 3–5, the testing time is 6 sec (see Plunkett et al., 2008)

visual map receives input vectors whose values freely vary. For the acoustic stimuli, each label is represented by a vector of dimension four (e.g. $[0,0,0.7,0.7]$). As with the infants, we have assessed the performance of the two architectures by measuring their response to the average and modal test stimuli. The measured response is a function of the quantisation error that we will define in the next section.

Network Architectures

We propose two neural network architectures that replicate the five experiments described in Plunkett et al. (2008). The two architectures use self-organising maps. They differ from each other in the number of maps involved. The first architecture can be seen as a more abstract implementation of the second. The first architecture contains a single map that receives both the visual and the acoustic stimuli. The second architecture contains three maps: a visual, an acoustic and a global map. The single map architecture implements the proposal that the auditory stimulus acts as a tag for the visual stimulus since it functions as a additional element to the visual feature vector. The three-map architecture permits some pre-processing of auditory and visual information before they are combined, implementing the idea that auditory and visual stimuli can stand in a more abstract relationship to each other.

Self-Organising Maps Self-organising maps consist of a set of units, spatially organised in regular grids. During the training phase, each map is presented with a set of input vectors that it learns to categorise. To each map unit u is associated a weight vector W_u of the same dimension as the input vectors. The i-th element of the weight vector W_u represents the weight connecting unit u to the i-th element of input vectors. The weight vectors are initialised to small random values. The learning algorithm behaves as follows: the input vectors are presented to the network in random order. After each presentation of a vector, its best matching unit is identified. This is the unit whose weight vector is closest to the input vector itself (in Euclidean distance). Next, the best matching unit's weights and the weights of its neighbouring units are adjusted towards the input vector, according to the equation:

$$W_u(t+1) = W_u(t) + s(u,t)a(t)(I(t) - W_u(t)) \qquad (1)$$

where $W_u(t+1)$ and $W_u(t)$ are the weight vectors associated to unit u at time $t+1$ and t respectively, $s(u,t)$ is the neighbourhood function associated with u at time t, $a(t)$ is the learning rate at time t, and $I(t)$ is the input vector presented to the network at time t. When u is the best matching unit for $I(t)$, we call the difference $\|I(t) - W_u(t)\|$ the 'quantisation error' of the network with respect to $I(t)$ (and we denote this error by $Q_e(t)$). In general, the value of the neighbourhood function shrinks over time, and the learning rate lowers over time. Commonly, for a self-organising map to learn to categorise a set of inputs, the inputs need to be presented to the network more than once. This does not seem very psychologically plausible. We will see that our model is an exception: each input needs to be presented only once to the network.

The Specificity of our Model According to Eq. 1, the weight update depends on the neighbourhood function and on the learning rate. In our case, the neighbourhood function entails that only the weights of the best matching unit are updated during the learning phase (whereas the weights of the neighbouring units remain unchanged). This is a simplification of our model that will be abandoned in future work.

The learning rate function is the most original part of the model. In contrast to the commonly adopted learning rate function, our function does not depend on time. Instead, it depends on the quantisation error, that is the Euclidean distance between the weight vector being updated and the input vector that produces this update: the higher the distance the higher the learning rate, the lower the distance the lower the learning rate. That the learning rate function depends on the quantisation error can be related to attention. The fact that the learning rate is higher when the input stimulus is new (and its quantisation error presumably high) could correspond to the observed phenomenon that infant attention is higher when exposed to novel stimuli. Furthermore, our learning rate is inversely proportional to the dimensionality of the input vectors. The fact that the learning rate is inversely proportional to the dimensionality of the input vectors derives from the fact that the same Euclidean distance is more significant for vectors with lower dimensionality than for vectors with higher dimensionality. From a cognitive point of view, this could reflect the fact that the increased cognitive load makes the input distances less significant: infants might be more sensitive to punctual distances when they have less information to process than when they have more. More precisely, our learning rate function is of the form $a(t) = \alpha e^{\beta Q_e(t)}$, where β is a scaling factor associated to each map, and α scales the learning rate according to the input dimensionality ($\alpha = 0.4/$ input dimensionality).

A final difference to the standard self-organising map architectures is that the model requires inputs to be presented to the network only once. We believe this contributes to the psychological plausibility of the model as this is the schedule in which the input stimuli are presented to infants in the five experiments that the model attempts to reproduce.

Network Architectures We propose two alternative architectures based on self-organising maps.

One-Map Architecture. This architecture consists of a single self-organising map, with 9 units, organised in a rectangular grid. To each unit is associated a weight vector of dimension eight, which is the length of the vectors that can be received as inputs by the map. The first four values of the weight vectors are weights to the visual inputs, the last four values are weights to the acoustic inputs. The weight vectors associated to each unit of the map are initialised to small random values. These values are uniformly distributed between the maximal and minimal values of the corresponding components of the input vectors, scaled by a factor of 0.3. The

neighbourhood function entails that only the best matching unit is affected by the learning algorithm. The inputs are presented to the network once, in random order. Weight update is performed by using the update rule described in Eq. 1, with the learning function $a(t)$ that is a function of the quantisation error, as described above, with $\beta = 2.3$.

Three-Map Architecture. In this architecture, we have three 9-unit maps: a visual, an acoustic, and a global map. The three sets of initial weight vectors are initialised as for the one-map architecture. The visual map receives as input the visual stimuli, whereas the acoustic map receives as input the acoustic stimuli. The global map receives as input the activity of the visual and acoustic maps in response to their respective input stimuli, where a map's activity is the collection of the activities of each unit, and a unit's activity is inversely proportional to the quantisation error ($e^{-Q_e(t)}$). For the visual and acoustic maps the learning rate function is the same as for the one-map architecture. In the case of the global map the learning rate function parameter β is 3.5.

Results

In order to replicate the five experiments reported in Plunkett et al. (2008), we trained both architectures (one-map and three-maps) by presenting once, in random order, the relevant input stimuli. For Exps. 3 & 5 that deal with acoustic as well as visual stimuli, we consider the whole architecture. In contrast, for Exps. 1 & 2, in which only the visual stimuli are relevant, we consider only the visual portion of the two architectures. In the one-map architecture, this is done by ignoring the weights connecting the global map to the acoustic stimuli. In the three-maps architecture, this is done by considering only the visual map.

We then assess the categorisation of the stimuli performed by the networks. Following Mareschal and French (2000) and Westermann and Marechal (2004) we adopt the network error as an analogue of looking time in infants. In our context, the error that we consider as analogous to infant looking time is the quantisation error. More precisely, we define the 'Network Looking Time' (*NLT*) as a monotonic function of its quantisation error ($NLT = 3Q_e^{0.2}$). We assume that if the network looking time corresponding to modal visual stimuli is systematically higher than the network looking time corresponding to the average visual stimulus, then the network exhibits a novelty preference for the modal stimuli and has, therefore, formed a single category. In contrast, if the network looking time corresponding to the average stimulus is systematically higher than the one corresponding to the modal stimuli, we will conclude that the network has formed two distinct categories. Since the test stimuli involve only visual stimuli, the acoustic pathways in the network are ignored during testing. All results refer to the performance of 24 networks per experiment. In the following, we catalogue the results for the five experiments. For each experiment, we consider the two architectures together. Table 2 summarises the one and three-map performance across the five experiments.

Experiment 1. We trained and tested the visual portion of 24 networks with the visual stimuli from the Broad Condition. Once the training phase was complete (8 cycles), we measured the network looking time for the average test stimulus and the average network looking time for the two modal test stimuli and computed their ratio. For both architectures, the results are shown in Table 2. In both cases, the network looking time for modal stimuli is higher than for the average test stimulus, thus indicating that the networks have formed a single category. This corresponds to infant behaviour in the Broad Condition, as reported in Table 1.

Experiment 2. Analogously to Exp. 1, networks were trained and tested using visual stimuli from the Narrow Condition. Results are shown in Table 2. In both cases, the network looking time for the average test stimulus is higher than for the modal test stimuli, thus indicating that the networks have formed two distinct categories. This corresponds to infant behaviour in the Narrow Condition .

Experiment 3. Networks were trained and tested with pairs of visual and acoustic stimuli (labels) in which a distinct label is associated with each visual category. In the one-map architecture, both the visual and the acoustic stimuli were simultaneously presented to the network. In the three-maps architecture, the visual and acoustic stimuli were first presented to the visual and acoustic maps, respectively. The acoustic and visual map activation was recorded and sent to the global map. Once the familiarisation phase was complete, the global map's performance was assessed by measuring the network looking time corresponding to the modal and average stimuli. In both architectures, the acoustic part of the networks was ignored during testing. Results are shown in Table 2. In both the one and three-map architectures, the network looking time for the average test stimulus is higher than for the modal test stimuli, thus indicating that the networks have formed two distinct categories. This result replicates the findings with infants that visual categorisation is not affected by the presence of labels, as long as these are correlated to the visual stimuli (Exp. 3 of Plunkett et al., 2008).

Experiment 4. We trained and tested the 24 networks as in Exp. 3, with the difference that labels were pseudo-randomly associated to visual stimuli (two visual stimuli from the same category were associated with the same label, the other two were associated with a different label). For both architectures, results are shown in Table 2. In both cases, there is no significant preference for the test stimuli, thus indicating that no category has been formed. Hence, labels can disrupt category formation when decorrelated with visual inputs. This corresponds to infant behaviour in Exp. 4 of Plunkett et al. (2008).

Experiment 5. Similarly to Exps. 3 & 4, we trained and tested the networks, this time with a single label consistent with all visual stimuli. The results reported in Table 2 show that in contrast to Exps. 3 & 4, the error corresponding to the modal stimuli is significantly higher than the error for the average stimulus, indicating that the map has formed a single

Experiment	Exp1	Exp2	Exp3	Exp4	Exp5
1-map					
LT average:modal	2.67:3.00	2.64:2.16	2.64:2.22	2.49:2.28	2.10:2.76
ratio LT (SD)	47.4% (1.5)	55.0% (0.5)	54.0% (0.65)	52% (6.3)	43.9% (5.2)
t-test (2-tailed)	t(23)=-10.56	t(23)=50.07	t(23)=31.64	t(23)=1.55	t(23)=-6.67
p-value	$\ll$ 0.0001	$\ll$ 0.0001	$\ll$ 0.0001	0.14	$\ll$ 0.0001
3-maps					
LT average:modal	2.55:2.94	2.64:2.34	2.61:2.07	2.40:2.25	2.19:2.52
ratio LT (SD)	46.6% (0.35)	53.0% (0.11)	56.0% (0.8)	51.7% (4.7)	46.5% (0.65)
t-test (2-tailed)	t(23)=-52.16	t(23)=126.74	t(23)=40.35	t(23)=1.71	t(23)=-26.51
p-value	$\ll$ 0.0001	$\ll$ 0.0001	$\ll$ 0.0001	0.10	$\ll$ 0.0001

Table 2: Network Looking Time during Testing

category. Hence, the labels have changed the way in which the visual stimuli would otherwise have been categorised in the absence of labels. Indeed, the same stimuli without labels gave rise to two distinct categories, as shown by Exp. 2. This corresponds to the results obtained with infants in Exp. 5 of Plunkett et al. (2008).

Behaviour during the Training Phase Plunkett et al. (2008) report infant looking time during the familiarisation phase. These findings show that infant looking time during the familiarisation phase is higher in Exp. 1 than in Exp. 2, and it is higher in Exps. 3–5 than in Exp. 2 (see Table 3).

	Infants	1-map	3-maps
Experiment 1	5.81 (1.56)	6.53 (0.19)	6.53 (0.23)
Experiment 2	4.34 (1.79)	5.83 (0.02)	6.07 (0.02)
Experiment 3	7.30 (1.74)	6.07 (0.02)	5.53 (0.37)
Experiment 4	7.44 (1.31)	6.51 (0.47)	6.23 (0.28)
Experiemnt 5	7.39 (1.23)	6.93 (0.30)	6.58 (0.33)

Table 3: Infant and network looking time during familiarisation.

To a great extent, the network looking time during familiarisation in our model follows the same pattern as infant looking time. Table 3 shows that for both architectures the network looking time during training is significantly higher in Exp. 1 than it is in Exp. 2 (p's $\ll$ 0.0001). When we compare infant looking time and network looking time in the familiarisation phase of Exp. 2 versus Exps. 3–5, we see that infant looking pattern is fully matched by the one-map architecture. Indeed, in the one map-architecture the network looking time during training is systematically higher in Exps. 3–5 than it is in Exp. 2 (p's $\ll$ 0.0001). This does not hold in the three-maps architecture.

The difference in looking time across the experiments in infants and in networks can be explained in a similar manner. Both in infants and in networks, the difference between Exps. 1 & 2 can be explained with the different average distance among the stimuli presented in the two experiments: this distance is usually higher in Exp. 1 than it is in Exp. 2. In contrast, the difference in looking time during Exp. 2 and Exps 3–5 in infants can be explained with the increased cognitive load deriving from the presence of labels in the last three experiments. In the one-map architecture the difference

can be explained in a similar way, with simple linear algebra: in Exps. 3–5 the dimension of the input vectors is higher than in Exp. 2. As a consequence, the errors (and therefore the Network's Looking Time) are higher in Exps 3–5 than in Exp. 2. The different pattern for the three-maps architecture depends on implementation details, related to the different range of input values to the visual and global maps. We conjecture that the same performance as for the one-map architecture can be achieved by normalising the range of values in the visual input patterns and in the visual and acoustic map activation patterns. The other mismatches between infant and network looking time during training will be the object of future work.

Discussion and Conclusion

We have provided a neural network model of the impact of labels on visual categorisation. In five sets of simulations we have reproduced Plunkett et al. (2008)'s laboratory experiments with infants showing that labels can affect the way in which perceptual stimuli are categorised. In Plunkett et al. (2008), infants' categorisation is assessed by measuring their looking time at test stimuli, following the well-established novelty preference procedure. Similarly, in this work we have assessed networks' category formation by measuring their looking time at test stimuli. In turn, network looking time is a function of the quantisation error.

For all five experiments, the model's looking preferences mimic those of infants. Hence, when shown a set of visual inputs from the so-called Broad Condition, the networks look longer at the modal visual stimuli than the average stimulus indicating that the network has formed a single category. In contrast, when shown visual inputs from the Narrow Condition during familiarisation, networks look longer at the average stimulus at test indicating that they have formed two categories. These results also mimic those of the infants in Younger's (1985) study and offer an alternative implementation of infant categorisation effects to those provided by (Mareschal & French, 2000) and (Westermann & Marechal, 2004). An advantage of the current implementation is that it requires only a single presentation of the familiarisation stimuli and it makes use of unsupervised, biologically plausible learning algorithms. The pattern of categorisation changes yet again when the same visual stimuli are accompanied by

labels: visual stimuli from the Narrow Condition produce two distinct categories just in the case that labels are correlated with visual category membership in the Narrow Condition. When labels are randomly associated with the visual stimuli, the networks show no systematic preference for either the average or modal stimuli indicating a disruption of the category formation process. Finally, the networks demonstrated a novelty preference for the modal stimuli when the accompanying label is the same for all familiarisation stimuli, indicating the formation of a single category as in the Broad condition (Exp. 1).

To a large extent, our model's behaviour also reproduces infant looking time during familiarisation. Notwithstanding the shortcomings of the three-map architecture, we have shown that this general approach to modelling the impact of labels on infant visual categorisation offers a powerful new tool for understanding the nature of the mechanisms that underlie infant looking preferences in novelty preference tasks. In particular, the model demonstrates a psychological plausibility in that it requires only single exposure to familiarisation stimuli in order to capture the wide range of looking preferences reported by Plunkett et al. (2008). The success of the one-map architecture in capturing the experimental finding adds credence to the view that infants use labels as tags to the visual stimuli. It is not necessary to assume that labels are invitations to form categories or that they highlight the commonalities between objects (Waxman & Markow, 1995). In the model, labels **are** the common feature that binds the objects together in a category.

The ability of the model to mimic infant's looking preferences in Exp. 5 is worthy of further examination. Recall that Exp. 5 exploits the Narrow condition visual stimuli together with a single common label during familiarisation. From a statistical learning perspective, this label is entirely redundant during familiarisation since it is a constant across all trials. In general, neural network models exploit the variability between input stimuli to organise their patterns of connectivity and internal representations. One might have predicted, therefore, that the networks in Exp. 5 would ignore the label since it seems to provide no distinguishing information across trials. Nevertheless, we have demonstrated that a single common label has a dramatic impact on the network's classification of the visual stimuli, causing it to treat them as a single category rather than two.

The intuition of the label being redundant in this experiment is nevertheless not entirely misleading. In fact, when the network in Exp. 5 is trained with standard parameters for multiple epochs, it learns to segregate the visual stimuli into two categories rather than one, thereby ignoring the label. In other words, the formation of the single category by the network in Exp. 5 is a transient effect. This finding has several important implications: First, it predicts that infants should show similar transient effects such that if they were continuously trained on the label-object contingencies of Exp. 5, infants would eventually form two visual categories. Second,

the transition from a single to a two category representation in the model implies that the label changes its status from being associated with a single visual category of objects to being associated with two discriminable visual categories of objects. This suggests that the model has the potential to represent a hierarchy of categorical organisation as a result of the introduction of a common label for members of the hierarchy. These transitions are emergent properties of a self-organising system which do not require explicit instruction or feedback. The model also predicts that the demonstrated impact of labels on categorisation in 10-month-old infants does not represent an end-point of learning but rather is a step *en route* to the development of a more structured system—perhaps a system that underpins the organisation of the mental lexicon itself.

References

Crick, F. (1989). The recent excitement about neural networks. *Nature*, *337*, 129–132.

Kohonen, T. (1997). Self-Organizing Maps, volume 30 of Springer Series in Information Sciences. *Springer, Berlin*.

Li, P., Zhao, X., & Whinney, B. M. (2007). Dynamic self-organization and lexical development. *Cognitive Science*, *31*, 581–612.

Mareschal, D., & French, R. (2000). Mechanisms of categorization in infancy. *Infancy*, *1*, 59–76.

Mayor, J., & Plunkett, K. (2008). Learning to Associate Object Categories and Label Categories: A Self-Organising Model. *Proceedings of the 30th Annual Cognitive Science Society. Austin, TX: Cognitive Science Society.*

Miikkulainen, R. (1997). Dyslexic and Category-Specific Aphasic Impairments in a Self-Organizing Feature Map Model of the Lexicon. *Brain and Language*, *59*(2), 334–366.

Mirolli, M., & Parisi, D. (2005). Language, as an Aid to Categorization: a Neural Network Model of Early Language Acquisition. *In A. Cangelosi et al., Proceedings of the 9th Neural Computation and Psychology Workshop, Singapore: World Scientific.*

Plunkett, K., Hu, J., & Cohen, L. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*, 665–681.

Plunkett, K., Sinha, C., Møller, M., & Strandsby, O. (1992). Symbol Grounding or the Emergence of Symbols? Vocabulary Growth in Children and a Connectionist Net. *Connection Science*, *4*, 59–76.

Waxman, S., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*, 257–302.

Westermann, G., & Marechal, D. (2004). From Parts to Wholes: Mechanisms of Development in Infant Visual Object Processing. *Infancy*, *5*(2), 131–151.

Younger, B. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, *56*, 1574–1583.