INFANCY, *5*(2), 199–207

THEMATIC COLLECTION: COMMENTARIES

How Do Simple Connectionist Networks Achieve a Shift From "Featural" to "Correlational" Processing in Categorization?

Michael S. C. Thomas
School of Psychology
Birkbeck College
University of London

Three developmental connectionist models simulate a purported shift from "featural" to "correlational" processing in infant categorization (models: Gureckis & Love, 2004/this issue; Shultz & Cohen, 2004/this issue; Westermann & Mareschal, 2004/this issue; empirical data: Cohen & Arthur, 2003; Younger, 1985; Younger & Cohen, 1986). In this article, the way in which the models are able to simulate the behavioral data is revealed, and their respective theoretical commitments are evaluated. Together the models argue that the shift from featural to correlational processing in infant categorization might be illusory, as these models are able to replicate the key behavioral features while processing correlations right from the start. As such, they claim the behavioral description of a shift is not reflected at the level of mechanism.

In this article, I establish how the three connectionist models of infant categorization presented by Westermann and Mareschal (2004/this issue), Shultz and Cohen (2004/this issue), and Gureckis and Love (2004/this issue) achieve a shift from ap-

Supplementary materials to this article are available on the World Wide Web at http://www.infancyarchives.com

Requests for reprints should be sent to Michael S. C. Thomas, School of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, United Kingdom. E-mail: m.thomas@psychology.bbk.ac.uk

pearing to process inputs featurally to processing them according to correlations between features. Such a shift has been proposed to account for the change in the categorization behavior of infants from 4 months to 10 months, assessed using a habituation paradigm (e.g., Younger & Cohen, 1986). In what follows, I take the empirical evidence for granted and concentrate on the properties of the models that allow them to simulate the empirical data.

First, however, a word about the modeling enterprise. It is important to recognize the role of the implemented models, such as those presented in these three articles, in advancing theory in developmental psychology. Implemented models demonstrate the viability of particular learning mechanisms in achieving the behavioral changes that we see across development. Although a model cannot on its own demonstrate that a particular theory is correct, modeling explores the patterns of behavioral data that can be generated by systems embodying particular principles of processing. As such, models serve the crucial role of widening the set of candidate inferences we can make from behavior to the processing structures that produce it (for a discussion, see Lewandowsky, 1993; Thomas & Karmiloff-Smith, 2002). Moreover, the process of model implementation forces a degree of detail in theory formation that often leads to an advance in our understanding of the target domain. In my view, each of the models discussed here is successful in enriching the theoretical debate concerning the nature of the representational changes that characterize infant categorization.

Somewhat unusually, in the current context we have three competing models that target the same empirical phenomenon: the shift from "featural" to "correlational" processing in infant categorization. Comparison of these models might prompt a wider metatheoretical discussion about the way in which one maps from the implementational details of each model to the theory that it embodies, and the way in which one evaluates respective models that are equally successful in simulating a given set of empirical data. Unfortunately, there is insufficient space for this kind of discussion here. Instead, I concentrate on three issues: (a) Why are the models successful in simulating the data? (b) What are the particular claims of the respective models? (c) What conclusions should we draw from the three models with regard to infant categorization?

HOW DO THE MODELS WORK?

First, then, let us focus on how the models work. Two factors are apparent when one inspects the empirical data that are simulated by these models (Cohen & Arthur, 2003; Younger & Cohen, 1986; see Figure 1 in Shultz & Cohen, 2004/this issue, depicting the results from Younger & Cohen, 1986, Experiment 2, and Figure 2 depicting the results from Younger & Cohen, 1986, Experiment 3 and Cohen &

Arthur, 2003). First, there is a general change in overall looking time to the test stimuli, with 10-month-olds exhibiting shorter looking times than 4-month-olds. Ten-month-olds seem less surprised. Second, the main evidence taken to indicate a shift from featural to correlational processing is a modulation of the response to the *uncorrelated* novel item, that is, a novel item that shares features with those introduced in the familiarization condition but that has these features in a novel configuration (so violating the correlational structure). When compared to the *correlated* item (an item either from the training set or with familiar features in a familiar configuration) and an item with *novel features,* the uncorrelated novel item is found to be relatively more surprising at 10 months than it is at 4 months.

One way to explain this modulation is to suppose that infants independently represent featural and correlational information and become increasingly sensitive to the correlational structure of the stimulus as they develop. The uncorrelated item therefore becomes more novel and surprising with age because it violates the correlational structure of the familiarization set.

However, an alternative explanation of the behavioral data is possible if one makes the following assumption. Initially the uncorrelated item is partially similar to the items in the familiarization set because it shares features with them. In this scenario, the infant has a threshold to determine which items fall within a known category (or categories) formed from the items in the familiarization set and which items fall outside as novel, prompting longer looking. Suppose that, at 4 months, this threshold is set low, so that not only the familiar correlated item but also the partially similar uncorrelated item exceed the threshold and are classified as "known" items and therefore uninteresting. The item with the novel features, however, fails to exceed the threshold, is classified as unknown, and so triggers additional looking. Now the shift to the 10-month-old pattern can be explained in two ways. Either there is a *change in the threshold,* so that the partially similar uncorrelated item no longer passes as known, or there is a *change in similarity,* so that the uncorrelated item no longer looks as similar to the items in the familiarization set.

If one assumes an initial continuum of similarity (correlated → uncorrelated → novel) it can be demonstrated that a simple mathematical model with three free parameters is sufficient to simulate the pattern of the empirical data depicted in Figures 1 and 2 in Shultz and Cohen's (2004/this issue) article. The three parameters are (a) the known-category threshold, (b) the relative similarity of test items, and (c) a base-rate level of surprise (for reference, this mathematical model can be found at www.infancyarchives.com). However, such a model would not serve as an explanation of the data because it does not specify any mechanism of change. It fails to explain why the category threshold or the similarity between test items should alter as a function of development. In contrast, all three articles instantiate precise claims concerning the way in which experience-dependent category-formation systems can lead to an apparent shift from featural to correlational processing across development.

In this article, I contend that the models presented by Westermann and Mareschal (2004/this issue), Shultz and Cohen (2004/this issue), and Gureckis and Love (2004/this issue) all achieve the featural to correlational shift (FC shift) by employing the threshold method, the similarity method, or a combination of both—rather than a shift in sensitivity between independent representations of featural and correlational information.

The mechanisms at play can be easily illustrated by considering a stripped-down version of the categorization problem and a highly simplified neural network model. A minimal model of the problem domain coded over four features can be constructed using a simple back-propagation network with only four input units, three hidden units, and four output units. The model demonstrates the FC shift across training (as marked by a modulation in response to the uncorrelated test item with age). Its internal representations can be readily examined. In the minimal model, it is apparent that training the network on the familiarization set causes an alteration both in the relative similarity of the internal representations of the test stimuli and of the thresholds that define the known category. (Details of the minimal model are available at www.infancyarchives.com, along with files to allow the reader to explore this model using the Tlearn network simulator; Plunkett & Elman, 1997.)

Were this a stripped-down model of infant categorization, its explanation of the FC shift would be that a 10-month-old undergoes more representational change during exposure to a familiarization set than does a 4-month-old (i.e., the "10-month-old" network gets more training during exposure to the familiarization set than the "4-month-old" network). More important, the minimal model highlights the key theoretical claim of the three models. A standard back-propagation network always functions by combining information. Each hidden unit is driven by the combined activation arriving from the input units. Each output unit is driven by the combined activation arriving from the hidden units. In the minimal model, it is apparent that although behaviorally the system might appear to show a shift in sensitivity from featural to correlational information, in fact at all times processing is driven by correlations between features. There is no initial stage of featural processing, just a change in the nature of correlational processing, with gradual shifts in thresholds and similarity patterns. The behavioral shift is simply not reflected at a level of mechanism.

The three models of infant categorization presented by Westermann and Mareschal (2004/this issue), Shultz and Cohen (2004/this issue), and Gureckis and Love (2004/this issue) all turn out to be more complicated versions of this minimal model. Each model introduces variations on the processing structure or the learning algorithm. However, the message is the same in each case. In terms of mechanism, the FC shift is illusory.

THE THREE MODELS

In the following sections, each model is assessed in terms of how it achieves the FC shift: by changing category thresholds or by changing representational similarity. Its additional theoretical commitments are then evaluated.

Westermann and Mareschal

Westermann and Mareschal (2004/this issue) propose a system that employs hidden units with tuneable Gaussian functions rather than the sigmoid processing units of standard back-propagation networks. In addition, the weights between the input units and hidden units are frozen at the beginning of training. Development from 4 months to 10 months is simulated by increasing the sharpness of the Gaussian functions in the hidden units. Westermann and Mareschal argue that training in pattern mode is important in their model. *Pattern mode* means that the model is trained on one pattern at a time to criterion. This contrasts with *batch mode,* where the network is repeatedly exposed to the whole training set. However, Westermann and Mareschal do not compare the performance of their model under these two conditions (pattern vs. batch) to demonstrate the strength of their claim.

In terms of the FC shift, the Westermann and Mareschal model opts for changing the similarity of the representations of test items. This is implemented by having a fuzzier (more overlapping) representation of the items at 4 months. At 10 months, the model is given sharper representations that decrease the similarity between the correlated and uncorrelated items in the test set. Output units do not have thresholds (simply summing the activation arriving from the hidden units) and so cannot modulate the responses. A change in similarity alone is sufficient to produce the FC shift.

Westermann and Mareschal justify the use of Gaussian units by appeal to receptive fields in visual processing. However, the decision to freeze the weights between input units and hidden units during training seems somewhat strange: Why should half of the system show activity-dependent changes but not the other half? In addition, the mechanism of development remains unexplained: What drives the change in similarity? The authors explicitly remain neutral about whether the change is maturational or activity dependent, but this aspect of the model nevertheless remains unspecified. Finally, although the authors claim that training in pattern mode is an essential theoretical commitment, the Shultz and Cohen (2004/this issue) model (and indeed the minimal model) manage to demonstrate the FC shift while training in batch mode. Westermann and Mareschal argue for the importance of pattern mode on two grounds. The first is that training in pattern mode is more psychologically plausible (the infant sees one item at a time). Second, elsewhere Mareschal and colleagues (e.g., see Mareschal, French, & Quinn, 2000; Mareschal, Quinn, & French, 2002) have been concerned with explaining additional empirical data revealing interference effects in infant categorization and have published models exploring the conditions under which interference should and should not occur. Interference effects are a risk of training in pattern mode but not in batch mode.

Shultz and Cohen

Shultz and Cohen (2004/this issue) employ a cascade correlation network that alters its architecture during learning (by adding hidden units). This type of model is referred to as *generative.* In the Shultz and Cohen model, development from 4 months to 10 months is simulated by altering the depth of learning between the two ages. At 10 months, the system is required to learn the training set more accurately. This requirement leads to the recruitment of more hidden units on average (4 months: 3.1; 10 months: 3.3) and more sweeps of training (56.2 vs. 59.1) to reach the requisite level of accuracy. Shultz and Cohen compare their generative model to a standard back-propagation network, which has an architecture that is fixed during training. The fixed architecture model does not demonstrate any FC shift. Shultz and Cohen therefore argue that qualitative shifts in behavior across development necessitate the postulation of systems that alter their architecture as a function of experience.

Although inspection of connection weights files would be necessary to be sure, it seems likely that the Shultz and Cohen model achieves the FC shift by both changes in threshold at output and changes in similarity in the hidden unit representations. The requirement for a greater number of hidden units in the 10-month-old condition implicates changing similarity, and the greater amount of training in the 10-month-old condition implicates larger weights and altered thresholds (this can be seen explicitly in the minimal model; see www.infancyarchives.com).

With regard to the claim that generative models are necessary to simulate the FC shift, this position seems weakened by the fact that Westermann and Mareschal's (2004/this issue) fixed architecture network appeared able to demonstrate it. In addition, the minimal model with a fixed $4 \times 3 \times 4$ architecture also exhibited a version of the FC shift. However, it is also true that the minimal model was unable to demonstrate the FC shift with only two hidden units. Therefore Shultz and Cohen are probably correct in arguing that the emergence of the FC shift in these models depends in part on the available representational resources and the particular encoding of the problem domain. Moreover, as Gureckis and Love (2004/this issue) also employed a type of generative model, this is clearly a viable approach to modeling the shift. Hopefully fixed architecture and generative models will generate different predictions and become empirically distinguishable in this domain.

Gureckis and Love

Gureckis and Love (2004/this issue) introduce a model that has elsewhere been applied to phenomena within adult categorization (SUSTAIN). The model includes parameters for the attentional biasing of features that turn out to be diagnostic for a category and for either supervised or unsupervised learning. In the current domain, the model is trained in an unsupervised mode. The model employs a self-organizing algorithm to derive a representation of the training set. The algorithm resembles cascade correlation in that as part of learning, the network can add additional representational resources (hidden units) to capture further distinctions in the problem domain. The internal representations are then used to reproduce the training items at output. The FC shift is simulated in two ways, either by altering a parameter that increases the recruitment of new hidden units during training (similar to Shultz & Cohen, 2004/this issue), or by reducing the amount of noise added to blur the encoding of stimuli (similar to Westermann & Mareschal, 2004/this issue). In the younger network the noise is higher, increasing the overlap between the representations of stimuli, whereas in the older network the noise reduces. Gureckis and Love claim that their model is parsimonious because it provides a mechanistic continuum from infant to adult categorization.

Both of the solutions proposed by Gureckis and Love appeal to changes in similarity to achieve the FC shift.

The parsimony of the model is an advantage. Such an advantage could also be claimed for any back-propagation network that simulated the FC shift, as these have also been applied to adult categorization (e.g., Gluck & Bower, 1988a, 1988b). However, with its additional parameters, the SUSTAIN model appears to be more powerful in its scope than simple back-propagation networks.

DISCUSSION

At first glance, the success of the three models is counterintuitive. Here is a set of infant data taken to imply a qualitative shift from featural processing to correlational processing. Here are three models that can demonstrate the behavioral shift without ever doing pure featural processing. In fact, the featural processing stage in these models tends to involve showing the uncorrelated item in such a way that (when looked at through scrunched-up eyes or some such representational manipulation) it looks pretty similar to the training items. This similarity then resolves with development (when you unscrunch your eyes or some such representational manipulation) so that the uncorrelated item does not look so similar to the training items any more. However, in all these models, right from the start, processing is correlational through and through. Each of the models proposes that some processing parameter has changed across development:

representational acuity for Westermann and Mareschal (2004/this issue), depth of learning for Shultz and Cohen (2004/this issue), and hidden unit recruitment or representational acuity for Gureckis and Love (2004/this issue). The minimal model offered another processing parameter: number of training epochs. Each model requires theoretical justification for its chosen manipulation to explain developmental change.

For theories of infant categorization, the take-home message from these three models is the following: There is no shift from featural to correlational processing at the level of mechanism. The qualitative shift applies only the description of behavior and is achieved by continuous (quantitatively) changing mechanisms.

However, the three models only serve to establish the viability of this hypothesis. Were there to exist some independent source of empirical evidence demonstrating that 4-month-olds possess independent representations of featural and correlational information, it might then be more parsimonious to adopt the alternate hypothesis—that the FC shift is achieved by an alteration in the relative sensitivity to these two types of information across development. One source of evidence might be if infants can decouple correlational information (or relational structure) from featural information to transfer it across situations. For instance, one might point to claims that infants can generalize the correlational structure of auditory tones across situations where the auditory tones themselves are different (Marcus, Vijayan, Bandi Rao, & Vishton, 1999). Weaker evidence might come from claims that older children are able to separate correlational from featural information in the context of analogical transfer and similarity judgments (Gentner, 1988). One might then argue that, because we have to postulate independent representations of featural and correlational information in the older child anyway, why not postulate them in the infant, too? However, it is also possible that each of these behavioral phenomena may also be amenable to simulation in models that do not separate correlational and featural information at the mechanistic level, just as we have seen with the current models (e.g., for such a claim regarding the Marcus et al., 1999, data, see Seidenberg & Elman, 1999).

The great advantage of the three articles is that they introduce detailed, implemented models with proven viability in capturing behavioral patterns. This contrasts with the alternative theory requiring independent representation of features and correlational structure, which has yet to be demonstrated as viable in an implemented learning mechanism. Given the choice, I would tend to place my money on a successful implementation rather than a verbal theory, but I have lost money before.

REFERENCES

Cohen, L. B., & Arthur, A. E. (2003). *The role of habituation in 10-month-olds' categorization*. Unpublished manuscript.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59,* 47–59.

Gluck, M., & Bower, G. (1988a). Evaluating an adaptive network model of human learning. *Journal of Memory and Language, 27,* 166–195.
Gluck, M., & Bower, G. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experiment Psychology: General, 117,* 225–244.
Gureckis, T. M., & Love, B. C. (2004). Common mechanisms in infant and adult category learning. *Infancy, 5,* 173–198.
Lewandowsky, S. (1993). The rewards and hazards of computer simulations. *Psychological Science, 4,* 236–243.
Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283,* 77–80.
Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in infancy. *Developmental Psychology, 36,* 635–645.
Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cognitive Science, 26,* 377–389.
Plunkett, K., & Elman, J. L. (1997). *Exercises in rethinking innateness: A handbook for connectionist simulations.* Cambridge, MA: MIT Press.
Seidenberg, M., & Elman, J. L. (1999). Do infants learn grammar with algebra or statistics? *Science, 284,* 433.
Shultz, T. R., & Cohen, L. B. (2004). Modeling age differences in infant category learning. *Infancy, 5,* 153–171.
Thomas, M. S. C., & Karmiloff-Smith, A. (2002). Residual normality: Friend or foe? *Behavioral and Brain Sciences, 25,* 772–780.
Westermann, G., & Mareschal, D. (2004). From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy, 5,* 131–151.
Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development, 56,* 1574–1583.
Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development, 57,* 803–815.