A Generative Neural Network Analysis of Conservation

Thomas R. Shultz
LNSC, Department of Psychology, McGill University
1205 Penfield Avenue
Montréal, Québec, Canada H3A 1B1
shultz@psych.mcgill.ca

Conservation

One of the most well studied phenomena in cognitive development is conservation. Conservation involves the belief in the continued equivalence of two physical quantities over a transformation that appears to alter one of them. An example of conservation presents a child with two identical rows of evenly spaced objects. Once the child agrees that the two rows have the same number of objects, the experimenter transforms one of the rows, e.g., by pushing its items closer together. Then the experimenter asks the child whether the two rows still have the same amount or whether one of them now has more. Piaget (1965) and other researchers found that children below about six years of age respond that one of the two rows, usually the longer row, now has more than the other. In contrast, children older than six years respond that the two rows still have equal amounts, i.e., they conserve the equivalence of the two amounts over the compressing transformation.

Despite the many empirical studies of conservation, the cognitive mechanisms underlying conservation acquisition remain obscure. One way to explore such cognitive mechanisms is with computer simulations in which the details of knowledge representations and processing mechanisms must be fully specified.

Cascade-correlation

A successful modeling algorithm for cognitive developmental phenomena is cascade-correlation. This is a generative algorithm for learning in feed-forward neural networks (Fahlman & Lebiere, 1990). It builds its own topology as it learns, by recruiting new hidden units into the network as it needs them. Such networks undergo not only quantitative adjustments in connection weights but also qualitative adjustments in network topology. There have been cascade-correlation models of balance scale phenomena, causal predictions of potency and resistance, seriation, integration of velocity, time, and distance cues, and acquisition of personal pronouns (Shultz, Schmidt, Buckingham, & Mareschal, 1995).

Simulations

Here I report on the simulation of five well known conservation phenomena with neural networks constructed by the cascade-correlation algorithm: (1) shift from nonconservation to conservation beliefs (acquisition effect), (2) emergence of correct conservation judgments for small quantities before larger quantities (problem size effect), (3) conservation of discrete quantities before continuous quantities (discrete advantage effect), (4) nonconservers' choice of the longer row as having more items than the shorter row (length bias effect), and (5) younger children conserving until they see the results of the transformation (screening effect).

Training

Networks were trained in an environment with very few constraints. Inputs described equivalence conservation problems in which rows of objects were described in terms of their perceptual characteristics, namely length and density. Target feedback supplied to the network concerned relative equality judgments comparing the two rows. Transformations included those that alter number (addition and subtraction) and those that preserve number (elongation and compression). Addition and subtraction transformations each altered a row by one item. Elongation and compression transformations decreased or increased the density of the row by one level, respectively. See Table 1 for some example transformations. In the standard rows, there were three levels of length, ranging from 2-6, and two levels of density, ranging from 2-4. Conservation experiments typically present only a few density levels but several levels of length. The quantities for numerical comparisons were computed as number = length x density. In this way, networks could learn about number from the perceptual characteristics of items arranged in rows with a constant within-row density.

Testing

A randomly selected 1/4 of the problems were excluded from training for use as test patterns. Most assessments were performed on test patterns rather than training patterns in order to insulate network performance from the particulars of training.

Acquisition

Networks learned the training problems and generalized well to the test problems, not merely memorizing the problems, but abstracting an underlying function.

Problem Size

Networks showed a problem size effect by performing better on problems in which the number of the smaller row was less than 9 than on problems in which the number of the smaller row was greater than 15. This was evident at all phases of training, except very early, where networks had not learned enough, and very late, where networks had reached a ceiling of performance on all problem sizes. Problem size effects are pervasive in human quantitative judgments; simulations suggest that they result from an analog representation of number.

Discrete Advantage

The discrete advantage effect was captured by adding small amounts of random noise to outputs in the training and test patterns for continuous quantities, which are considered difficult to estimate accurately. It took longer for networks to learn with noisy than with noiseless outputs. Also, except early and late in training, networks performed worse on noisy than on noiseless problems. Thus, the networks were sensitive to noise, but not fatally so. The absence of a noise effect at the beginning of training reflects inability to solve either type of problem. The disappearance of the noise effect at the end of training reflects ceiling levels of performance.

Length Bias

The networks showed biases like those observed in children. They learned to use the dominant dimension of length, and then noticed the compensating dimension of density, before correctly integrating the two dimensions.

The initial length bias was due to learning that longer rows often have more items than shorter rows, particularly in addition and subtraction transformations when density is held constant. This explanation is consistent with the idea that very young children do not show a length bias because they have not yet learned that length is a correlate of number (Miller, Grabowski, & Heldmeyer, 1973). Length bias did not occur in an alternate environment in which length, rather than density, was held constant during transformations. These bias results underscore the tension between perception and cognition in conservation tasks. What the child knows (e.g., that a transformation does not change a quantity) appears to conflict with what she sees (e.g., that one row is longer, and thus seems more numerous than the other).

Screening

The screening effect refers to young children conserving only until they see the results of a transformation (Miller & Heldmeyer, 1975). As long as the effects of the transformation are screened from view, they conserve, but when the screen is removed, they revert to nonconservation. This was simulated by removing information about the appearance of the transformed row after it was transformed, causing more conservation early in training.

Network Analysis

To determine the roles of particular hidden units, information critical to perceptual and cognitive solutions was deleted from the test problems. Missing input critical to a perceptual solution involved the length and density of the post-transformation row, whereas missing input critical to a cognitive solution involved the nature of the transformation. Analysis of errors caused by these deletions indicated that most hidden units played a role in either perceptual or cognitive solutions and a few of them played a role in both solution types.

Conclusions

These simulations captured a variety of effects in the conservation literature and supported the correlation-learning explanation of length bias. They achieved better and more comprehensive coverage of natural conservation phenomena than have previous simulations.

Acknowledgments

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. Adam Waese, Yasser Hashmi, Yuriko Oshima-Takane, and Sylvain Sirois commented on earlier drafts.

References

Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2* (pp. 524-532). Los Altos, CA: Morgan Kaufmann.

Miller, P. H., Grabowski, T. L., & Heldmeyer, K. H. (1973). The role of stimulus dimensions in the conservation of substance. *Child Development, 44*, 646-650.

Miller, P. H., & Heldmeyer, K. H. (1975). Perceptual information in conservation: Effects of screening. *Child Development, 46*, 588-592.

Piaget, J. (1965). *The child's conception of number*. New York: Norton.

Shultz, T. R., Schmidt, W. C., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive development with a generative connectionist algorithm. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 205-261). Hillsdale, NJ: Erlbaum.

Transformation	Length	Density	Row
Pre-transformation	2	2	o o o o
Add	2.5	2	o o o o o
Subtract	1.5	2	o o o
Elongate	4	1	o o o o
Compress	1.33	3	o o o o

Table 1: Example transformations.