A Connectionist Model of the Development of Transitivity

Thomas R. Shultz (thomas.shultz@mcgill.ca)
Department of Psychology, McGill University, 1205 Penfield Avenue
Montreal, QC H3A 1B1 Canada

Abbie Vogel (abbie.vogel@mail.mcgill.ca)
Department of Psychology, McGill University, 1205 Penfield Avenue
Montreal, QC H3A 1B1 Canada

Abstract

A modular connectionist model covers all six established phenomena in transitivity development in children and predicts a new effect. In contrast, a symbolic-rule hypothesis based on logic captures none of these effects and is directly contradicted by one of them. In the model a constraint-satisfaction network generates a response based on input from a feed-forward comparison module and the particular question asked. Cycles to saturate the response module implement response times.

Psychology of Transitivity

Piaget and his colleagues (Inhelder & Piaget, 1964; Piaget, 1969) designed the transitivity problem to assess the development of children's logical-inference abilities. This problem often employs sticks (or times) of different length, as shown in Figure 1. Given, for example, that a child learns that stick 2 is longer than stick 1, and that stick 3 is longer than stick 2, can the child infer that stick 3 must be longer than stick 1? This is not a perceptual problem in that the child only identifies a stick by its unique color, never seeing the actual stick lengths. Piaget's evidence suggested that correct untrained inferences, such as comparing sticks 1 and 3, did not emerge until around seven years of age, thus providing an index of the child's entry into the stage of concrete operations.

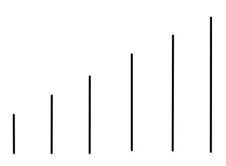

Figure 1: A six-stick version of the transitivity task.

Piaget's long-dominant view of the transitivity task as being solved by logic was ultimately contradicted in an experiment measuring the time it took for people of different ages to make various inferences (Trabasso, Riley, & Wilson, 1975). Using a six-item version of the task like that in Figure 1, Trabasso et al. trained 6-year-olds, 9-year-olds, and university students on all adjacent pairs of sticks and then asked about all possible pairs of sticks, varying the question between *Which stick is longer*? and *Which stick is shorter*? Five different effects were reported.

1. A serial-position effect: learning the adjacent pairs near the ends of the array before the pairs near the middle.
2. A distance effect: faster inferences about pairs that are farther apart in length than for pairs close together in length.
3. An anchor effect: faster inferences about pairs involving an end anchor (sticks 1 or 6) than for pairs not involving an end anchor.
4. A congruity effect: faster inferences when the term used in the question (e.g., longer) is compatible with an end anchor (e.g., the longest stick) in the pair being compared than when the question term (e.g., longer) is incompatible with an end anchor in the pair being compared (e.g., the shortest stick).
5. An age effect: older participants learned the adjacent pairs faster and made inference comparisons faster and more accurately than did younger participants.
6. Other experiments with different comparison tasks found that the distance effect diminished with increasing age (Duncan & McFarland, 1980; Sekuler & Mierkiewicz, 1977).

The first four of these effects have been replicated in a wide range of tasks involving symbolic comparisons along a dimension, e.g., numerical comparisons (Banks, 1977; Duncan & McFarland, 1980; Leth-Steenson & Marley, 2000; Sekuler & Mierkiewicz, 1977).

The distance effect was particularly damaging to Piaget's logical-inference interpretation because it is precisely opposite to what Piaget would presumably predict. Assuming that each inference takes some constant time, Piaget would have to predict that the more inferences required to make a comparison, the longer the comparison would take. For example, comparing sticks 2 and 3 requires no inference at all because participants are trained on such adjacent pairs. In contrast, comparing sticks 2 and 4 requires a single inference from two premises ($S2 < S3$ and $S3 < S4$, therefore $S2 < S4$). And comparing sticks 2 and 5 requires two inferences (the previous inference plus this one: $S2 < S4$ and $S4 < S5$, therefore $S2 < S5$). The larger the split (or difference) between sticks, the more inferences would be required. The splits are conventionally termed 1, 2, and 3 in these three comparisons, respectively.

Because of the distance effect in their response-time data, Trabasso et al. concluded that people don't use logical inference per se to solve this task. They argued instead that participants construct a spatial image of the sticks while being trained on adjacent pairs and then consult this image when asked to make another comparison. The farther the sticks are apart within this spatial image, the easier it is to make a correct comparison. Despite a recent resurgence of interest in studying and modeling transitivity, there has been no computational model that covers all six of these effects. It is, in fact, computationally unclear how the brain might construct and consult spatial images in this way.

The purpose of the present work is to build such a model with cascade-correlation (CC), a neural-network learning algorithm that has been used to simulate many other phenomena in cognitive development (Shultz, 2003). Another reason to use CC is that it searches in topology space, building the network, as well as in weight space.

Like other feed-forward neural algorithms, CC produces responses in more or less constant time, and thus is not naturally suitable for covering response-time effects. To add this capability, we used a modular system of two networks, which we call constraint-satisfaction cascade-correlation (CSCC). A CC network learned to judge the relative lengths of adjacent sticks and a constraint-satisfaction (CS) network used that information plus information contained in the question to generate a response. The number of update cycles that the response module required to settle into a steady state was taken as an index of response time.

Method

The CSCC modular network system is shown schematically in Figure 2.

Comparison Module

The CC comparison module is on the left side of Figure 2. Inputs to the comparison module describe the colors of the two sticks being compared and were coded in a binary *n*th-unit fashion (1 for the color of a stick and 0s elsewhere for colors that are not involved). The 12 inputs (the same 6 colors for each of two sticks) were fully connected to a single output unit having a sigmoid activation function, which coded a length comparison with targets of -0.5 if the left (L) stick is longer and 0.5 if the right (R) stick is longer. Comparison networks were trained on all ten adjacent pairs of sticks until all output values were within score threshold of their targets on all of these ten training pairs. Order of the two sticks being compared is counterbalanced across comparisons.

We implemented age differences by using different values of score threshold: 0.5 for adults, 0.55 for 9-year-olds, and 0.6 for 6-year-olds. This is consistent with finding that older people learn more from the same experiences than young children do (Case, Kurland, & Goldberg, 1982), a principle that has been successfully used to simulate age differences in learning in several other CC simulations (Shultz, 2003). Different score thresholds would also work for capturing developmental effects here, but these particular values produced overall proportions correct that were very close to those reported by Trabasso et al. (1975) for their different age groups. We ran 12 networks at each of the three score-threshold levels, matching the *n*s at each age level in the Trabasso et al. experiment. Full details of the CC algorithm are discussed elsewhere (Shultz, 2003).

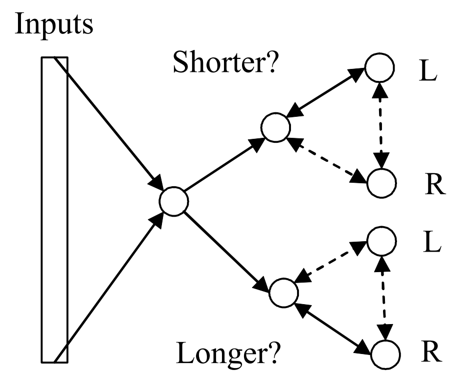

Figure 2: CSCC modular networks for transitivity.

Response Module

After training, output from the comparison module served as input to a three-unit response network. Activation on the other two units in this CS network represented the left or right sticks as being the correct response to the question being asked. As shown on the right side of Figure 2, the precise form of this response network varied according to the question being asked. Recall that the target output of the comparison network was 0.5 when the right stick was longer, and -0.5 when the left stick was longer. Consequently, if the question was *Which stick is longer*?, then there were positive weights (0.5, signified by a solid line) between the comparison unit and the right (R) unit and negative weights (-0.5, signified by a dashed line) between the comparison unit and the left (L) unit. If the question was *Which stick is shorter*?, then the signs of these weights were reversed; there were positive weights between the comparison unit and the left unit and negative weights between the comparison unit and the right unit. The basic principle underlying these weight settings is to enhance the activation value of the side unit corresponding to the stick that is longer when the question term is *longer*, and to enhance the activation value of the side unit corresponding to the stick that is shorter when the question term is *shorter*. More generally the idea is to activate the correct response and inhibit an incorrect response.

Connections between the left and right units were always negative to reflect the idea that these two units are competing with each other. Unlike the comparison unit, these two side units had no external inputs; all of their input came from inside the response network.

As is typical with CS networks, weights in this response module were bidirectional, with one weight going in each direction between any two units. As in other CS simulations (Kunda & Thagard, 1996; Shultz & Lepper, 1996), we assume here that these networks are constructed on the fly by participants in response to their particular experimental setting and the question being posed to them. There is no assumption that participants are conscious of this construction. It is rather that the design of a response module is strongly constrained by the participant's understanding of the experimental situation and question.

All three units in a response module started out with an initial activation value of 0. At every cycle, three units were randomly selected, with replacement, to have their activations updated. In each such update, net input to the updated unit i was computed as:

$$net_i = in\left(\sum_j w_{ij} a_j\right) + ex(input_i) \qquad (1)$$

where a_j is the activation of each sending unit j, w_{ij} is the relevant connection weight, $input_i$ is any external input to the receiving unit, and *in* and *ex* are parameters scaling influences internal or external, respectively, to the network. These last two parameters were both set to 0.1 in our simulations, but a wide range of values work equally well.

If this net input was positive, it was added to the receiving unit's current activation $a_i(t)$ after being scaled by the distance of that current activation from the activation ceiling of 1.0:

$$a_i(t+1) = a_i(t) + net_i(ceiling - a_i(t)) \qquad (2)$$

Alternatively, if the net input was negative, it was added to the receiving unit's current activation $a_i(t)$ after being scaled by the distance of that current activation from the activation floor of -1.0:

$$a_i(t+1) = a_i(t) + net_i(a_i(t) - floor) \qquad (3)$$

An overall measure of the degree to which a CS network has settled into a stable state is its *goodness*, computed as the sum of triple products of unit activation values and the relevant connection weight plus the sum of the products of external inputs and activation values:

$$goodness = \sum_{ij} w_{ij} a_i a_j + \sum_i input_i a_i \qquad (4)$$

Equations 1-4 are fairly standard in the CS-network literature (Shultz, 2001). In this kind of scheme, goodness rises as units have their activations updated and eventually levels off as activations stop changing. Examples are provided in Figure 3 in terms of goodness changes over update cycles for networks with three different levels of comparison inputs. We identified the cycle at which goodness starts to reach asymptote as no goodness change greater than .02 (asymptote-threshold parameter) for 8 consecutive cycles (asymptote-patience parameter). These parameter values were selected because they correspond to our visual impressions of when goodness values approach asymptote. A range of different threshold and patience values works equally well, although sufficiently extreme parameter values can blur differences between sticks and between conditions. Figure 3 shows that networks settle quicker with higher, and more decisive, comparison activations.

To cover the congruity effect, we multiplied comparison inputs by 0.8 whenever there was an anchor stick that was incompatible with the term used in the question. This is a computational shortcut consistent with the idea that the congruity effect is based on semantic interference between incompatible terms, some of which have to be translated to a compatible form to answer the question (Banks, 1977). When asymptote was reached, the comparison unit, left or right, with the higher activation was taken as the response module's answer to the question that was posed.

Before activation-update cycles began, all the connection weights and external inputs had their initial values randomized a bit by adding or subtracting up to 10% of their initial values in a uniform distribution. This is to reflect the fact that not all participants interpret the experimental procedures and questions in exactly the same way. A wide variety of randomization values work equally well to implement such individual differences.

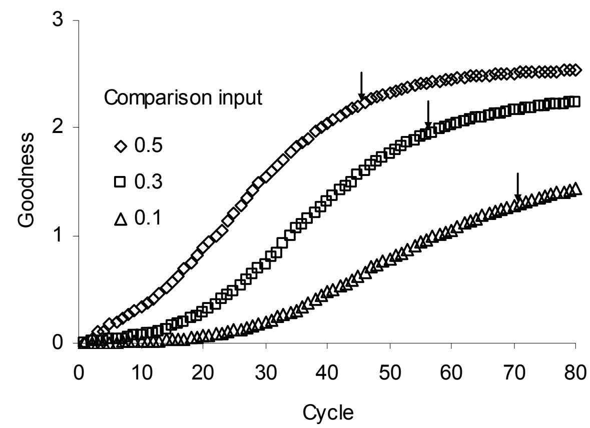

Figure 3: Increasing goodness over activation-update cycles in a CS response network at three levels of comparison input. Arrows indicate the cycle at which a goodness asymptote was reached.

Results

Learning

Although CC networks are capable of recruiting new hidden units if they are needed, none of our networks did so. This indicates that the problem of learning *n*th-unit, binary-coded, adjacent stimulus pairs is a rather simple linearly-separable problem. The mean number of epochs taken to learn the training patterns was 7.2 for score threshold of 0.6, 7.7 for score threshold of 0.55, and 10.2 for score threshold of 0.5.

The serial-position effect for training is evident in Figure 4. A score-threshold x training-pair mixed ANOVA of comparison-network error yielded a quadratic trend for training pair, $F(1, 33) = 279$, $p < .001$. With no interaction with score-threshold, this shows the serial position effect at each age: better learning of training pairs at the ends of the array than in the middle. There was also a main effect of score-threshold, $F(2, 33) = 19.59$, $p < .001$, capturing the superiority of older, deeper learners.

Inference

The theoretically-important distance effect is shown in Figure 5. In a score-threshold x split ANOVA of cycles to settle, the largest effect is a linear trend for split, $F(1, 33) = 582$, $p < .001$, confirming a strong distance effect at every score threshold, representing the three different ages. Network responses were faster the larger the split between the sticks being compared. It is also evident that the distance effect diminished a bit with decreasing score threshold, representing increasing age.

In Figure 6, a score-threshold x end-anchor ANOVA of cycles to settle reveals a main effect of anchor, $F(1, 33) = 166$, $p < .001$, simulating the finding that performance is quicker when an end anchor is present. A score-threshold x end-anchor interaction, $F(2, 33) = 6.73$, $p < .01$, predicts that the anchor effect may also diminish with increasing age.

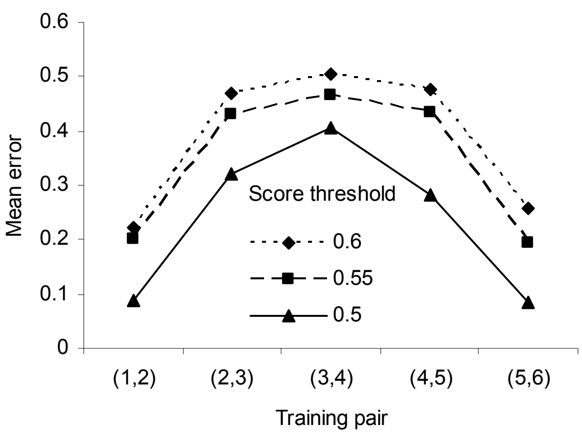

Figure 4: The serial-position effect: mean error for different training pairs and score thresholds.

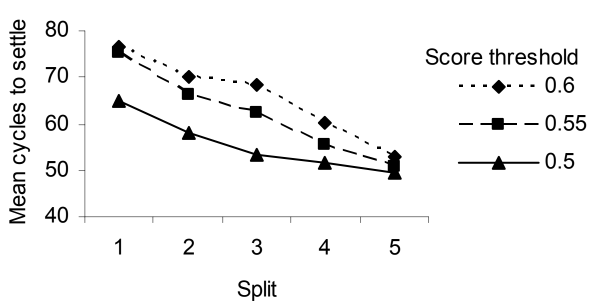

Figure 5: The distance effect: mean cycles to settle for different splits and score thresholds.

The congruity effect is plotted in Figure 7, in the form of an end-anchor x question interaction, $F(1, 33) = 288$, $p < .001$. This shows faster responding when there is compatibility between question and end anchor.

Knowledge-representation Analysis

Mean weights across 12 networks at each score-threshold level are plotted in Figure 8 in order to understand the knowledge representations acquired by the comparison networks. Recall that target output activation is negative when the left stick is longer, and positive when the right stick is longer.

Correct performance by a network can be understood by considering a few example weights. A large positive weight from the R6 input ensures positive output compared to any shorter comparison stick L1 to L5. A somewhat less positive weight from the R5 input produces positive output except when compared to the longer stick L6, in which case the stronger negative weight from the L6 input produces a negative output, signaling that the left stick is longer.

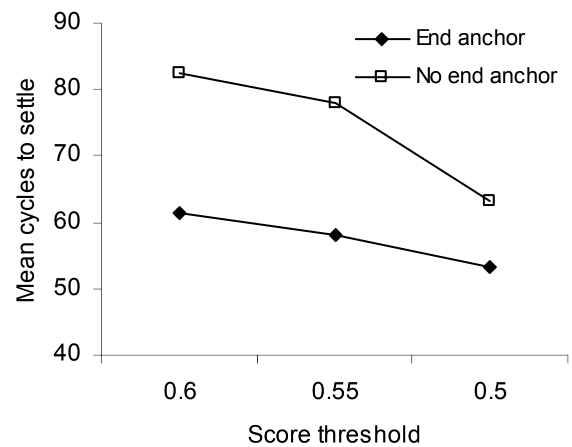

Figure 6: The anchor effect: mean cycles to settle for different score thresholds and the presence of an end anchor.

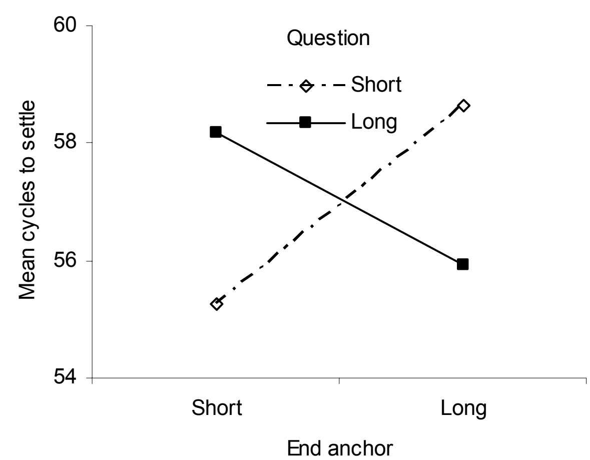

Figure 7. The congruity effect: mean cycles to settle for different size anchors and question phrasing.

The overall pattern of weights is V-shaped, seen most clearly at the lowest score-threshold of 0.5, representing adults. For the right sticks, weights are larger with increasing stick size; for the left sticks, weights are smaller with increasing stick size because the target output is negative when a left stick is longer. Fairly precise left-right symmetry in weight values on each branch of the V is important to enable accurate judgment of pairs that are close together (e.g., L2 vs. R3).

Serial-position Effect The fact that connection weights have a steeper slope near the ends of the array than in the middle explains the serial-position effect. More distinctive weights produce larger absolute comparison outputs, which are closer to their target values, yielding less network error.

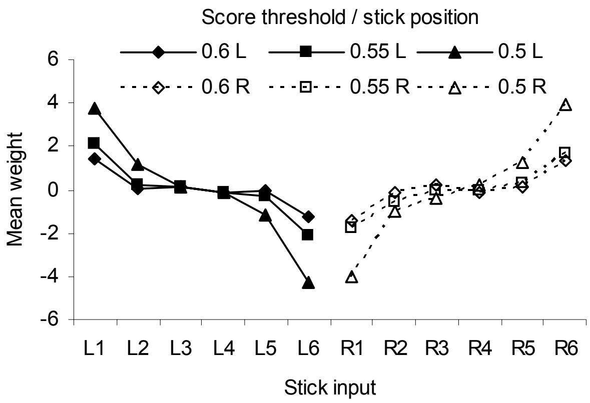

Figure 8: Mean weights in comparison networks from various input units at different score thresholds.

Distance Effect The manner in which this knowledge representation produces the distance effect is evident from the weight plot. Sticks close in size are more likely to have similar weights, producing small absolute comparison values, thus requiring more cycles to reach asymptote in the response module. In contrast, sticks that are farther apart in size are more likely to have larger differences in weight values, producing larger absolute comparison outputs, and thus requiring fewer response-module cycles.

Anchor Effect The manner in which this knowledge representation produces the anchor effect is also evident. Weights for the end anchor sticks (1, 6) have more extreme values than do weights for the other sticks, ensuring larger absolute comparison values and thus quicker responses when end-anchor sticks are involved in a comparison.

Developmental Effects The origin of developmental effects is also apparent from these knowledge representations. The lowest score-threshold of 0.5 (representing adults) produces the steepest V shape with the most easily distinguishable weights. The higher score-thresholds of 0.55 and 0.6 (representing 9- and 6-year-olds, respectively) produce progressively shallower V shapes with weights that are closer in size. The less distinctive the weights, the smaller the absolute output of the comparison module and the more cycles required to reach asymptote in the response module.

Discussion

Our model is a hybrid modular system, with a feed-forward CC network making a length comparison and a CS network using this comparison information along with question information to generate a response. This CSCC model simulated all of the established psychological effects in the development of transitivity in humans. Captured phenomena include the serial-position, distance, anchor, congruity, age-related improvement, and diminishing distance effects.

All of these effects followed naturally from the modular-networks model without any parameter tweaking or special manipulation of training patterns. In general, these effects were produced by the comparison network's natural tendency to learn to order the stimuli by length on its connection weights. The serial-position and anchor effects were due to the fact that these weights were more distinct near the ends of the array than in the middle. The distance effect arose from the fact that the relevant connection weights (those with non-zero inputs) were more distinct with sticks of more distinct size. The congruity effect arose from incompatibility between an anchor and the term used in the questioning, which was made to cause a small degradation of the comparison signal. The age-related improvement and diminishing distance effects were simulated by the familiar phenomenon of older individuals learning the problem more deeply than younger ones do.

Similar interpretations of the serial-position, distance, and anchor effects have been offered by other connectionist modelers (Leth-Steenson & Marley, 2000). But ours is the first model to capture all six effects and to offer novel connectionist interpretations of the congruity and developmental effects. Together these models show how transitivity phenomena can be explained in a neural fashion. We plan to review all of the recent simulations of human and animal data on transitivity in a fuller publication. Different models typically focus on somewhat different phenomena.

Does our model confirm Trabasso et al.'s (1975) hypothesis that people consult a visual image of an ordered spatial array of sticks to answer inference questions? The knowledge representation learned by comparison-module networks certainly does order the array of sticks by length. This learning is based merely on information about the relative lengths of adjacent pairs, without any information on how long the sticks actually are. Whether this knowledge representation constitutes a *visual* image, either in artificial networks or in real brains, is debatable. One way to investigate this issue in real brains might be to see if visual cortex becomes particularly active in brain images of people learning and solving transitivity tasks (Behrmann, Kosslyn, & Jeannerod, 1996). In any case, the simulation presented here demonstrates a fully specified functional account of transitivity development, whether assumed to be located in visual cortex, hippocampus, or other brain regions.

In contrast to these neural-network simulations, Piaget's original logical-inference view cannot account for any of these transitivity phenomena. Indeed its predictions for the effect of distance on response time are precisely the opposite of what actually occurs. Because Piaget's hypothesis can be naturally framed in terms of a recursively-applied symbolic transitivity rule, this issue can be viewed as another instance of the symbolic rules vs. subsymbolic connections debate that has dominated cognitive science for the past 18 years. As far as psychological development is concerned, results have consistently favored the connectionist approach because it typically covers a wider range of phenomena in a more principled fashion than does the symbolic rule-based approach (Shultz, 2003).

In a more detailed publication, we will present data and analyses of correct inferences by our networks. In general these data also mirror the performance of participants in Trabasso et al.'s (1975) experiment. There is no problem with speed-accuracy tradeoffs in our simulations because response times and errors are positively related.

Although it is beyond the scope of this paper to fully evaluate the various alternative psychological theories of transitivity, these theories do not seem capable of accounting for all the phenomena treated here.

Results indicated that age effects on transitivity tasks can reflect rather small quantitative differences in depth of learning, rather than major qualitative differences in type of processing. Similar results have been found in simulations of a number of other developmental phenomena, including seriation (Mareschal & Shultz, 1999), discrimination shift learning (Sirois & Shultz, 1998), and concept learning (Shultz & Cohen, 2004). Our explanation represents a radical departure from previous interpretations of these phenomena, which have tended to suggest that older children are doing something qualitatively different than are younger children. Because of its capacity for network growth, the CC algorithm is particularly well suited to discovering whether qualitative changes are necessary for capturing developmental change. Some developmental phenomena require such qualitative growth, while other developmental phenomena do not (Shultz, 2003). In addition, our model predicted a diminishing anchor effect with increasing age that could be tested with children.

To capture the heretofore elusive congruity effect we implemented a (shortcut) neural version of the idea that semantic incompatibility between an anchor and question term can slow performance. We also plan to implement an alternative hypothesis based on the notion that particular combinations of question and items serve to bias the participant's response at the start of a random walk towards one or another decision boundary (Link, 1990). Our current response module, with its random selection of units to update, might be adapted to implement the basic features of Link's hypothesis. Comparing results across the two techniques could indicate which hypothesis provides a better explanation.

Acknowledgments

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to the first author. We thank J. P. Thivierge, Frédéric Dandurand, and Yuriko Oshima-Takane for helpful comments on an earlier draft.

References

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11). New York: Academic Press.

Behrmann, M., Kosslyn, S. M., & Jeannerod, M. (Eds.) (1996). *The neuropsychology of mental imagery*. New York: Pergamon.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Duncan, E. M., & McFarland, C. E. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory and Cognition, 8*, 612-622.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation*. London: Routledge and Kegan Paul.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel constraint-satisfaction theory. *Psychological Review, 103*, 284-308.

Leth-Steenson, C., & Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review, 107*, 62-100.

Link, S. W. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology, 34*, 2-41.

Mareschal, D., & Shultz, T. R. (1999). Development of children's seriation: A connectionist approach. *Connection Science, 11*, 149-186.

Piaget, J. (1969). *The child's conception of time*. London: Routledge and Kegan Paul.

Sekuler, H., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48*, 630-633.

Shultz, T. R. (2001). Constraint satisfaction models. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 4). Oxford: Pergamon.

Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge: MIT Press.

Shultz, T.R., & Cohen, L. B. (2004). Modeling age differences in infant category learning. *Infancy*, *5*, 153-171.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review, 103*, 219-240.

Sirois, S., & Shultz, T. R. (1998). Neural network modeling of developmental effects in discrimination shifts. J*ournal of Experimental Child Psychology, 71*, 235-274.

Trabasso, T., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R. Falmagne (Ed.), *Psychological studies of logic and its development*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.