.

A Connectionist Model of the Development of Velocity, Time, and Distance Concepts

David Buckingham
Department of Psychology
McGill University
1205 Penfield Avenue
Montreal, Quebec, Canada H3A 1B1
dave@lima.psych.mcgill.ca

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

Abstract

Connectionist simulations of children's acquisition of velocity (v), time (t), and distance (d) concepts were conducted using a generative algorithm, cascade-correlation (Fahlman & Lebiere, 1990). Diagnosis of network rules were consistent with the developmental course of children's concepts (Wilkening, 1981, 1982) and predicted some new stages as well. Networks integrated the defining dimensions of the concepts first by identity rules (e.g., v = d), then additive rules (e.g., v = d-t), and finally multiplicative rules (e.g., v = d/t). Psychological effects of differential memory demands were also simulated. It is argued that cascade-correlation implements an explicit mechanism of developmental change involving incremental learning and qualitative increases in representational power.

Introduction

In classical physics, velocity is defined as v = d/t, time as t = d/v, and distance as d = vt. A number of developmental psychologists, beginning with Piaget (1969, 1970), have assessed the development of these three interrelated concepts in children. Much of the psychological research has been criticized for presenting children with problems that could be solved non-inferentially, by merely ignoring irrelevant information and perceiving the information that is being asked about (Levin, 1977; Wilkening, 1981). For example, when asked to compare the velocity of two toy trains, the child can ignore time and distance information and report on the perceived velocities.

Wilkening (1981) designed more valid tasks in which children were asked to infer velocity, time, or distance given information about the two defining dimensions. For example, in a distance-inference task, children were shown an apparatus that had, at one end of a footbridge, a dog and several other animals. The children were told that the other animals would run along the bridge as soon as the dog began to bark and would stop when the barking ceased. The task involved determining how far each animal would run. Thus, the children were given the characteristic velocity of the animals and the time they ran (the duration of barking), and asked to infer the distance they would run.

Wilkening studied the performance of three age groups: 5-year-olds, 10-year-olds, and adults. The findings included the following: (1) In the distance-inference task, all age groups used the correct multiplication rule, $d = vt$; (2) in a time-inference task, 10-year-olds and adults employed the correct division rule, $t = d/v$, whereas 5-year-olds used a subtraction rule, $t = d-v$; (3) in a velocity-inference task, the two older age groups used a subtraction rule, $v = d-t$, and the 5-year-olds used an identity rule, $v = d$.

Wilkening hesitated to draw strong conclusions concerning the concurrent development of the three concepts since it appeared that the tasks had differing memory demands. For example, in the distance-inference task, subjects of all age groups used an eye-movement strategy in which they followed the imaginary course of the animal as it ran across the footbridge. In terms of the time-inference task, Wilkening pointed out that information about both defining dimensions was available at the time of judgment. However, in the velocity-inference task, the subjects had to retrieve the time information from memory.

In a follow-up study, Wilkening (1982) attempted to increase the memory demands of the distance task by presenting time information (barking) before velocity information (animal identity) and lessen the memory demands of the velocity task by visually presenting the time information. The modifications partially supported his hypothesis in that 5-year-olds were observed to use an additive rule ($d = t+v$) in the distance task. However, the results of the velocity task remained unchanged. Thus, it remains to be seen if the mastery of time before velocity concepts is an accurate description of the developmental course or a memory artifact of Wilkening's tasks.

Cascade-correlation

Here we report on connectionist simulations of the acquisition of velocity, time, and distance concepts in an attempt to gain some explanatory insight into these psychological results. Our simulations employ cascade-correlation (Fahlman & Lebiere, 1990), a generative connectionist algorithm that begins with a minimal network topology determined solely by the number of input and output units. The input units, including an obligatory bias unit, are connected directly to the output layer. Thus, the initial network topology is like that of a perceptron. Training is carried out in a two-phase cycle. In the *output training phase*, the weights from input units and any installed hidden units are adjusted using first-order and an approximation of second-order error derivatives to minimize the total sum of squared error. When the error stops decreasing, the *input training phase* begins. Weights from input units to a pool of candidate hidden units are adjusted in order to maximize the correlation between candidate hidden units and the output error. The hidden unit with the activation that is most correlated with the output error is then installed into the network and the output training phase recommences.

Cascade-correlation has been used successfully to model children's performance on other developmental tasks including the balance scale (Shultz & Schmidt, 1991; Shultz, Mareschal, & Schmidt, in press), seriation (Mareschal & Shultz, 1993), the effects of potency and resistance on the magnitude of a physical effect (Shultz *et al*., in press), and personal pronouns (Shultz, Buckingham, & Oshima-Takane, in press). Several of these simulations involved rule-based stages, even though rules are not explicitly represented in the networks. For the domain of velocity, time, and distance, it was expected that the increasing non-linear computational power of a network that recruits hidden units as needed might provide insights into how children progress from simple identity rules to more complex multiplicative rule-like performance.

Simulations

The task of the networks was the same as for Wilkening's subjects. The networks had to predict, as output, the value of one dimension (e.g., velocity) given information about the other dimensions (e.g., distance and time). The initial network topology consisted of three input banks, one each for distance, time, and velocity information, connected to a single linear output unit. A linear output unit was used because it was the most natural way of producing a quantitative output similar to the responses made by subjects in Wilkening's experiments.

Input and Output Coding

Inference patterns were encoded as follows. Two input banks received dimensional values ranging from 1 to 5. The third bank received an input of 0 indicating that it was the dimension to be predicted. Each input bank had five input units for a total of 15 input units. A dimensional value *n* was encoded by assigning an activation of 1 to the *nth* input unit of the bank and 0 to all other units in the bank. We call this *nth* encoding. Thus, for a given inference pattern, one input bank would receive activations of 0 on all of its five input units indicating it was unknown. One unit of each of the other two input banks received an activation of 1. The remaining units in these banks received activations of 0. Notice that in *nth* encoding, the inputs initially possess cardinality but not ordinality (McClelland, in press); the network must learn the ordinal relations among the input values as it learns the velocity, time, and distance problems.

Target values for the output unit were calculated using the three Newtonian equations ($v = d/t$, $t = d/v$, and $d = vt$) respectively. In addition, distance target values were scaled by dividing by five so that the range would be identical to the target values of time and velocity inference patterns. Twenty-five instances of each of the three inference problem types were obtained by crossing the five levels of velocity, time, and distance for a total of 75 inference patterns.

Two sets of simulations are reported. The *accurate memory condition* simulations represent a situation in which memory demands across the three concepts are equal and minimal. The *limited memory condition* explores the effects of inaccurate memory for time information in velocity-inference tasks. Here, it was assumed that the likelihood of correct recall followed a normal distribution in that values closer to the actual time value would be more likely to be recalled than distant values. At each epoch of training, the time dimension of a given velocity inference problem was altered so that in general there was a 34%, 13%, 2%, and less than 1% probability that the time input value "recalled" by the network would be 1, 2, 3, or 4 integers away respectively from the actual value. In the remaining instances, the actual value was used as input.

An additional constraint was that the modified time input fall within the same range (1 to 5) used in the training set. Without such a restriction, networks in the limited memory condition would require more input units than those in the accurate memory condition. Therefore, if the selection of an integer either above or below the actual value was possible, a random choice was made among the two equally distant values. For example, if the actual input value was 3 there was a 34% chance that the network would receive 2 or 4 as input and a 13% chance that it would receive 1 or 5. Otherwise, 3 was used. If the input value was 5, then the chance of a 4, 3, 2, or 1 being "recalled" was 34%, 13%, 2%, and 1% respectively.

Only the time dimension inputs of velocity inference patterns were selected according to the criteria above. Target values were unaffected by this process. They were determined by dividing the distance dimensional value by the true time dimensional value.

Training and Rule Diagnosis

At each epoch of training, all 75 inference problems were presented to the network. Thirty networks in each condition were trained for a maximum of 1500 epochs. Every fifth epoch, the networks were diagnosed for rule use.

To compare network results with human performance, we diagnosed rules that best captured network performance on each problem type. We computed correlations between the network's responses and those predicted by various plausible rules such as identity ($v = d$, or $v = t$), addition ($v = d+t$, or $v = d-t$), or multiplication ($v = dt$, $v = t/d$, or $v = d/t$) rules. To capture consistent network performance, a given rule had to correlate positively with network responses, account for more than 50% of the variance in network responses (i.e., $r^2 > .50$), and account for more variance than other plausible rules.

Results

Accurate Memory Condition

A stage by epoch plot of a typical network in the accurate

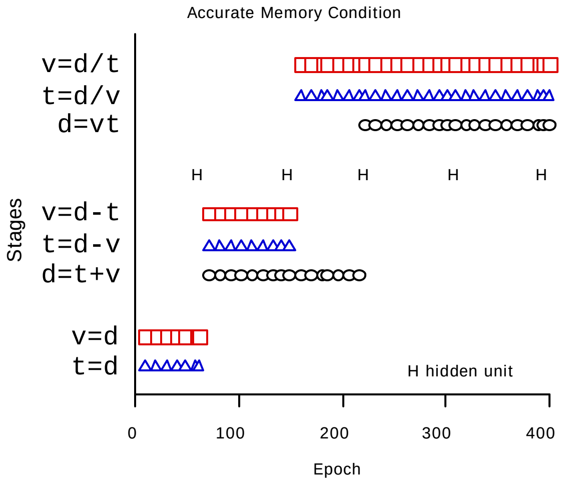

Figure 1: Stages achieved by one network in accurate memory condition, plotted every 10 epochs.

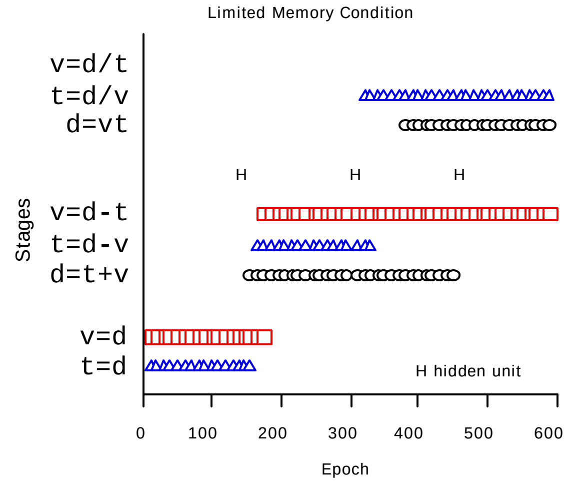

Figure 2: Stages achieved by one network in limited memory condition, plotted every 10 epochs.

memory[1] condition is shown in Figure 1. All 30 networks exhibited a qualitatively similar developmental course. As can be seen, time and velocity identity stages ($t = d$ and $v = d$) emerged early in training prior to the recruitment of a hidden unit. The mean epochs of onset and length of the identity stages are reported in the left half of Table 1. On average, these identity stages began together typically after 5 or 10 epochs of training and lasted for approximately 55 epochs. To assess how well particular rules captured network performance, we computed the mean across networks of the maximum r^2 value attained during each stage. During the identity stages, the mean was over 90% for both the time and the velocity identity rules, suggesting that both were good predictors of time and velocity inferences respectively. During this same period, the networks' responses to distance inference patterns were not captured by any of the rules that were tested.

Table 1: Mean Epoch of Onset and Length of Identity Stages.

	Condition			
	Accurate Memory		Limited Memory	
Stage	*M*	*SD*	*M*	*SD*
t=d				
onset	6.67	2.40	7.50	2.54
length	55.60	5.40	147.23	15.01
v=d				
onset	6.67	2.40	10.33	4.54
length	55.43	4.53	204.40	76.25

n = 30

Note. d = distance; t = time; v = velocity.

Following the recruitment of the first hidden unit, distance, time, and velocity inferences were captured by the additive rules $d = t+v$, $t = d-v$, and $v = d-t$ respectively. The mean epochs of onset and the length of these additive stages are reported in the left half of Table 2. Although all three began at approximately the 65th or 70th epoch, the distance additive stage lasted almost 90 epochs more on average than either of the other two stages. On average a maximum of over 80% of the variance in the three types of inferences was accounted for by the additive rules.

Table 2: Mean Epoch of Onset and Length of Additive Stages.

	Condition			
	Accurate Memory		Limited Memory	
Stage	*M*	*SD*	*M*	*SD*
d=t+v				
onset	70.93	3.83	155.86*	5.10
length	164.77	78.40	317.46	159.70
t=d-v				
onset	67.77	4.85	163.57	15.26
length	77.90	7.87	156.13	46.19
v=d-t				
onset	67.77	4.48	224.62**	79.87
length	77.93	7.04	437.17***	330.06

n = 30; * *n* = 28; ** *n* = 29; *** *n* = 18

Note. d = distance; t = time; v = velocity.

Multiplicative stages of time ($t = d/v$) and velocity ($v = d/t$) inferences began after 150 epochs of training following the recruitment of the second hidden unit. On average, the multiplicative stage of distance inferences ($d = vt$) began approximately 100 epochs later following the recruitment of the third or fourth hidden unit. The mean epochs of onset of the multiplicative stages are reported in the left half of Table 3. All three defining multiplicative rules of the stages eventually reached a maximum r^2 of 1.00. This occurred earlier for time and velocity inference patterns than for distance inferences.

[1] A preliminary report of the accurate memory condition is presented in Shultz *et al*. (in press).

Table 3: Mean Epoch of Onset of Multiplicative Stages.

	Condition			
	Accurate Memory		Limited Memory	
Stage	*M*	*SD*	*M*	*SD*
d=vt	248.97	83.94	519.80	214.29
t=d/v	150.60	8.26	330.93	50.94
v=d/t	150.27	8.52	711.32*	403.45

n = 30; * *n* = 19

Note. d = distance; t = time; v = velocity.

Limited Memory Condition

Eleven of the networks in the limited memory condition demonstrated a qualitatively similar developmental course to that depicted in Figure 2; all but one of the remaining networks followed a similar course except that the multiplicative stage of velocity inferences was eventually attained.

As in the accurate memory condition, for all 30 networks the time and velocity identity stages emerged early in training at approximately the same epoch, prior to the recruitment of a hidden unit. However, unlike in the accurate memory condition, the identity stage of velocity inferences lasted almost 60 epochs longer on average. The mean epochs of onset of the identity stages are reported in the right half of Table 1. Again the identity rules were good approximations of the predictions, accounting for an average maximum of over 90% of the variance.

For the majority of networks, the additive stages of distance and time inferences emerged soon after the recruitment of the first hidden unit. On average, the additive stage of velocity inferences emerged approximately 60 epochs later. Occasionally two hidden units were recruited prior to the onset of the velocity additive stage. Two networks skipped the additive distance stage and one skipped the velocity stage. The mean epochs of onset and the length of these additive stages are reported in the right half of Table 2. The mean length of the velocity additive stage reported in Table 2 is based on 18 networks since 11 networks did not progress beyond this stage and one, as mentioned, skipped the additive stage. Distance inferences based on the additive rule typically lasted longer than additive time inferences whereas additive velocity inferences were the most persistent. As in the accurate memory condition, a maximum of over 80% of the variance in the three types of inferences was accounted for by the additive rules.

The most striking difference between the accurate memory and limited memory conditions was that 11 limited memory networks did not progress to the multiplicative stage of velocity inferences. The network in Figure 2 depicts this result. All 30 networks progressed to the multiplicative stage of time and distance inferences. On average, the multiplicative stage of time inferences emerged after a second hidden unit was installed, about 190 epochs before the multiplicative stage of distance inferences. The distance multiplicative stage followed the recruitment of 3 or 4 hidden units. Of the 19 networks that progressed to the velocity multiplicative stage, on average they did so approximately 190 epochs after the distance multiplicative stage had been attained. The mean epochs of onset of the multiplicative stages are reported in the right half of Table 3. The defining multiplicative rules of the distance and time stages eventually reached a maximum r^2 of over 0.98. In contrast, for those networks that attained the velocity multiplicative stage, the defining rule reached a maximum r^2 of approximately 85%.

Hinton Analysis

It was apparent that hidden unit recruitment was necessary for the onset of additive and multiplicative stages given the abrupt transition to stages following the installation of a hidden unit.

In order to explicate the role of hidden units, Hinton diagrams of the incoming weights were drawn following the recruitment of hidden units. A representative example, taken from one network in the accurate memory condition, is shown in Figure 3. The two rows in the diagram represent the weights from the sending units to the output unit and first hidden unit respectively. The size of the weight corresponds to the size of the square; the color indicates the sign of the weight (white and black indicate positive and negative respectively). As can be seen, the weights from the time and velocity input banks (squares 6-10 and 11-15 respectively) to the hidden unit are of the same sign and opposite in sign to the weights from the distance input bank (squares 1-5). Thus, when a distance inference pattern was presented, the time and velocity inputs augmented each other. This gave rise to predictions that were correlated with the additive rule $d = t+v$. In contrast, when a time or velocity inference pattern was presented, the distance input would be counteracted by the velocity or time input. This gave rise to time and velocity inferences that correlated best with performance in which one dimension was subtracted from the other (e.g., $t = d-v$). Because of network complexities, Hinton analysis of the second and third hidden units were less revealing. However, given the relatively abrupt transition after the installation of these hidden units to multiplicative stages, the need for increased non-linearity seems evident.

Discussion

Simulation results for the most part matched those of Wilkening (1981, 1982) with humans. For distance inferences, there was a progression from an additive rule to the correct multiplication rule. For time and velocity inferences, networks began with an identity rule, progressed to an additive rule, and then finished with the correct multiplicative rule. Wilkening's human subjects did the same, except that they showed no identity rule for time inferences and failed to reach the correct multiplicative rule for velocity inferences.

The results of the limited memory condition suggest that memory demands may have been a factor in Wilkening's studies. By increasing the memory demands of the velocity

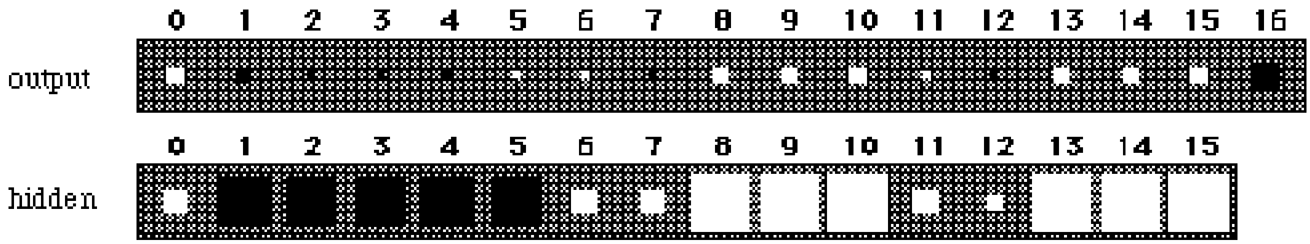

Figure 3: Hinton diagram of incoming weights after the recruitment of the first hidden unit for one network. Numbers refer to weights from bias (0), distance (1-5), time (6-10), velocity (11-15) and hidden (16) units.

task, velocity was delayed such that it was possible for a network to be making velocity inferences based on the identity rule and time inferences based on additive rules. Later when time inferences were based on the multiplicative rule, velocity inferences were typically additive. Finally, a number of limited memory networks did not attain the multiplicative stage of velocity inferences. Many of those that did regressed to the additive stage, some permanently.

The simulations suggest that, when all else is held constant, the developmental course is more consistent across concepts than Wilkening's results indicated. Identity stages emerge early for both time and velocity concepts followed by the additive stages of all three concepts and then the onset of multiplicative concepts for velocity and time prior to the eventual attainment of the distance multiplicative stage. We propose that the mechanisms of weight adjustment and hidden unit recruitment compounded with a learning environment in which all three concepts are being acquired constrains network development along these lines.

Identity stages emerge due to a combination of the limited processing ability of the initial perceptron-like architecture and the fact that one network is performing all three inference tasks. The initial network is not able to encode both the inverse relationship between time (velocity) and distance when making velocity (time) inferences *and* the direct relationship between time and velocity when making distance inferences. The algorithm is unable to find a set of weights for a perceptron to accommodate both roles of the time and velocity input. Therefore, the relationship of time (or velocity) information to the output error of distance and velocity (or time) inferences is obscured. In contrast, the relationship of the distance information to the output error is more clear since the role of distance is the same in either time or velocity inference problems. When the algorithm attempts to reduce the error across all three problem types, weight adjustment may be primarily influenced by the relationship of distance input to the error. When presented with a time (velocity) inference pattern, distance information has more influence than velocity (time) information. In contrast, when presented with a distance inference pattern, neither time nor velocity information has greater influence and the weights fail to encode the direct relationship between time and velocity information. Thus, identity stages emerge for time and velocity inference problems. However, neither identity, additive, nor multiplicative rules are able to capture the role of time and velocity with respect to distance inference problems during this period.

A similar argument could be put forward with respect to children. Early on, the child is confused about the inconsistent effects of time and velocity and focuses attention on the more consistent distance information when making velocity or time inferences. With respect to distance inferences, the child is at a loss as to how to solve the problem and may choose time or velocity depending on their salience.

Additive stages of each concept typically emerged after the installation of the first hidden unit. The network used this hidden unit to differentiate distance from time and velocity information by assigning one sign to the weights from the distance inputs and the opposite sign to weights from time and velocity inputs. Thus, the first hidden unit is able to encode the dual nature of time and velocity, at least in a simplistic manner, and as a result, the additive stages emerge.

The additive stages of all three concepts eventually were replaced by multiplicative stages. Typically, the time and velocity multiplicative stages emerged first followed by the distance multiplicative stage. One reason why the distance additive stage may have lasted longer than either the time or velocity additive stages was that a larger proportion of error was reduced during the distance additive stage than in the other two. This in turn delayed the onset of the distance multiplicative stage. Thus, the distance additive rule would seem to be a very good approximation of distance inference patterns. It may be that for people, use of an additive rule persists as a heuristic approach that is generally good enough.

Thus, cascade-correlation provides an explicit mechanism of transition in network performance -- weight adjustment and hidden unit recruitment. In human developmental terms we believe this to be akin to incremental learning in combination with increasingly non-linear representational abilities.

The simulation results, together with Wilkening's studies, make several predictions. First, the results suggest that if 5-year-olds can integrate distance *and* velocity and time *and* velocity information in an additive manner to infer time and distance respectively, they should be able to integrate distance and time information additively to make velocity inferences if the memory demands are minimized.

Second, whereas Wilkening's observation that 5-year-olds use a velocity identity rule is likely related to extraneous task demands, the simulations predict that children younger than five years of age make velocity inferences by focusing on distance information because of processing limitations. Moreover, although Wilkening did not find evidence of a time identity rule, our simulations suggest that younger children would use such a rule to make time inferences. Finally, these same-aged children would be expected to solve distance inference problems based on either time or velocity information depending on which is more salient. Therefore, future research should include younger children and manipulate the salience of velocity and time information in the distance inference task.

Third, if the task demands of velocity inferences were reduced, it is predicted that 10-year-olds would make velocity inferences by integrating the dimensions with the correct multiplicative rule. Therefore, the inability of Wilkening's subjects to correctly integrate time and distance information is again likely due to extra memory demands.

Finally, the simulations suggest that the distance multiplicative stage emerges after both time and velocity multiplicative stages. Since Wilkening did not study 10-year-olds' performance when an eye-movement strategy was not possible, it would be necessary to re-examine 10-year-olds under this condition.

In conclusion, the results of our simulations suggest that limited processing capacities of children lead to performance that is first characterized by identity rules for time and velocity inferences. Later, additive inference rules characterize performance. Finally, multiplicative rules characterize the integration of distance, time, and velocity inferences. Once again the cascade-correlation algorithm has demonstrated the ability to capture the emergence of rule-based stages in cognitive development.

Acknowledgments

This research was supported in part by a fellowship from FCAR Quebec to Buckingham and an operating grant from the Natural Sciences and Engineering Research Council of Canada to Shultz. We also would like to thank Scott Fahlman for providing the cascade-correlation code and Chris Schunn for providing code for automatic drawing of Hinton diagrams.

References

Fahlman, S. E. & Lebiere, C. (1990). The cascade-correlation learning architecture. Technical Report, CMU-CS-90-100, School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA.

Levin, I. (1977). The development of time concepts in young children: Reasoning about duration. *Child Development*, 48, 435-444.

Mareschal, D. & Shultz, T. R. (1993). A connectionist model of the development of seriation. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 676-681. Hillsdale, NJ: Lawrence Erlbaum.

McClelland, J. L. (In press). A connectionist perspective on knowledge and development. In G. Halford & T. Simon (Eds.), *Developing cognitive competence: New approaches to process modeling*. Hillsdale, NJ: Erlbaum.

Piaget, J. (1969). *The child's conception of time* (A.J. Pomerans, Trans.). London: Routledge & Kegen Paul. (Original work published 1946)

Piaget, J. (1970). *The child's conception of movement and speed* (G.E.T. Holloway & M.J. Mackenzie, Trans.). London: Routledge & Kegen Paul. (Original work published 1946)

Shultz, T. R., Buckingham, D., & Oshima-Takane, Y. (in press). A connectionist model of the learning of personal pronouns in English. In S.J. Hanson, M. Kearns, T. Petsche, & R.L. Rivest (Eds.), *Computational learning theory and natural learning systems*. Cambridge, MA: MIT Press.

Shultz, T. R., Mareschal, D., & Schmidt, W. C. (in press). Modeling cognitive development on balance scale phenomena. *Machine Learning*.

Shultz, T. R., & Schmidt, W. C. (1991). A Cascade-Correlation model of balance scale phenomena. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 635-640). Hillsdale, NJ: Erlbaum.

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

Wilkening, F. (1981). Integrating velocity, time, and distance information: A developmental study. *Cognitive Psychology*, **13**, 231-247.

Wilkening, F. (1982). Children's knowledge about time, distance, and velocity interrelations. In W.J. Friedman (Ed.), *The developmental psychology of time* (pp. 87-112). NY: Academic Press.