Denis Mareschal and
Michael S.C. Thomas
Centre for Brain and Cognitive Development School of Psychology
Birkbeck University of London

How Computational Models Help Explain the Origins of Reasoning

Abstract: Developmental psychology is ready to blossom into a modern science that focuses on causal mechanistic explanations of development rather than just describing and classifying the skills that children show at different ages. Computational models of cognitive development are formal systems that track the changes in information processing taking place as a behavior is acquired. Models are generally implemented as psychologically constrained computer simulations that learn tasks such as reasoning, categorization, and language. Their principal use is as tools for exploring mechanisms of transition (development) from one level of competence to the next during the course of cognitive development. They have been used to probe questions such as the extent of 'pre-programmed' or innate knowledge that exists in the infant mind, and how the sophistication of reasoning can increase with age and experience.

I. Understanding the Origins of Reasoning

The Swiss psychologist Jean Piaget was possibly the first man to ask how thinking emerged from the simple reflexes of the newborn to the abstract logical reasoning of the adult (Piaget 1971; Boden 1995). He saw himself as an empirical philosopher whose goal it was to answer the fundamental questions of epistemology (the study of the origins of knowledge) through rigorous experimentation. He asked how knowledge, especially abstract conceptual knowledge and logic-based reasoning, could emerge from a child's interactions with the world. Piaget produced a vast body of work exploring the development of concepts like Space, Time, Number, and Causality. He is widely recognized as having identified the key questions that have set the agenda for cognitive development research over the last 70 years.

Piaget was greatly influenced in his thinking by the philosophies of Kant and Bergson, but also by the Cybernetics movement of the early twentieth century. He believed that children constructed an understanding of the world through *active* engagement with the world, and that feedback on one's actions played a crucial role in learning and development. Unfortunately, he failed to ground many of his theoretical proposals because he lacked an appropriate vocabulary with which to express his dynamic and mechanistic ideas (Boden 1988). The arrival of computational modeling has provided a suite of powerful conceptual tools for addressing many of Piaget's original ideas.

1556-603X/06/$20.00©2006IEEE

Contemporary theories of cognitive development lie broadly along two distinct (albeit related) dimensions. One of these is the Nativist vs. Empiricist dimension (sometimes known as the *nature-nurture* debate). Radical Nativists believe that almost all knowledge is available to the infant prior to any experience. Learning only serves to fill in minor details. Radical Empiricists believe that the infant is born with powerful learning abilities but no prior knowledge. All knowledge is acquired through some form of experience with the world. Computationally, the approaches differ in the constraints that are placed on learning. In a *cognitive* computational model, the researcher uses psychological empirical data to constrain choices about representations and training sets. For example, a cognitive model designed to learn grammar might include input representations of individual words and output representations of who-did-what-to-whom. It might then be trained on samples of child-directed speech. Psychological approaches with a *nativist* leaning will produce models with tight constraints on their architectures, activation dynamics, input/output representations, or learning algorithms, so that the system can only support a restricted set of input-output functions. Exposure to a training set just serves to push the system into one of this limited set of states. For example, in the Chomskian theory of language acquisition, mere exposure to language is held to 'trigger' the selection of the correct subset of a Universal Grammar that is present at birth (see for example, Buttery 2004). Internal functions would be limited to tree-structures and operations over tree-structures. The latent structure of child-directed speech would have little causal role in shaping the final internal function. By contrast, models of a more *empiricist* bent will have fewer constraints on the functions that can be supported, so that the information in the training set plays a stronger role in determining which function is acquired. For example, some theories of the development of the visual system argue that when a fairly general self-organizing learning system is exposed to natural visual scenes, the latent statistical structure of these scenes is sufficient to generate many of the kinds of representational primitives observed in the low-level visual system, such as center-surround receptive fields (Field 1999a, b).A second dimension that distinguishes contemporary theories of cognitive development is the distinction between *symbolic* vs. *sub-symbolic* representations. Those in the *symbolic* camp believe that cognition is best characterized as a rule-governed physical symbol system. In this view, cognitive development consists in the modification of mental rules. By contrast, those in the *sub-symbolic* camp see cognition as a highly interactive dynamic system (e.g., an artificial neural network). In this system, the causal entities are continuous, distributed and cycling patterns of activation. Such networks do not operate as physical symbol systems, or at best approximate them in certain narrow circumstances. In this latter view, development consists in the continuous tuning of the underlying parameters of the cognitive system.

Piaget concentrated on studying development in normal populations. His theory therefore aimed to characterize the cognitive stages through which the 'average' child passes. However, in contemporary psychological theory, normal cognitive development is increasingly seen within the context of the ways that development can go wrong or operate sub-optimally. For example, children with genetic disorders can exhibit uneven cognitive profiles and sometimes learning disabilities, such as in Down syndrome, autism, and Williams syndrome. Apparently more circumscribed disorders can be found that differentially impact on language development or on the acquisition of reading, observed in Specific Language Impairment or dyslexia. Even within the normal population, children of the same age can differ in their cognitive ability. At the upper end, children are viewed as gifted, while at the extreme lower end, normal development begins to resemble disability. These variations in development throw into relief the boundary conditions that must shape normal development (Karmiloff-Smith 1998). In recent years, computational models of development have provided a productive tool to investigate how variations in the properties of learning systems—or the potentially enriched or impoverished learning environments to which they are exposed—can help or hinder the acquisition of cognitive abilities (Thomas and Karmiloff-Smith 2003).

In the rest of this article, we present a number of models to illustrate different aspects of cognitive development where computational approaches have produced material advances in our understanding of the origins of knowledge. As will become apparent in this review, most *symbolic* models have emphasized the tractability of the knowledge representations involved in cognitive development at the expense of implementing explicit transition mechanisms. In contrast most *sub-symbolic* models have emphasized the specification of a developmental mechanism at the expense of the tractability of the knowledge representations. In other words, to the extent that the human mind needs complicated and densely structured mental representations to deliver a cognitive skill, it becomes hard to see how these representations are acquired. The theorist is left with three ways out of this conundrum. Either complex behavior are generated by representations that are in large part innate; or we don't yet understand the full repertoire of learning mechanisms available to the human mind; or we are currently overestimating the complexity of the representations that the human mind needs to generate its complex behaviors.

II. Why Build Computational Models of Cognitive Development

A. The Computer Modeling Methodology

Computer models are invaluable tools for transforming developmental psychology from a descriptive science into a mature explanatory science (Mareschal and Thomas in press). When a researcher has to translate his or her underlying theory into an explicit computer model, he or she must now specify precisely what is meant by the various terms. Terms such as representations, symbols, and variables must have an exact definition to permit implementation. The degree of precision required to construct a working computer model avoids the possibility of arguments arising from the misunderstanding of imprecise verbal theories. For example, 'short-term memory' conveniently summarizes a cluster of human behaviors, yet it is another thing to build the persistence of information over time into a processing system. How much information is stored? How long is it stored for? There is no longer any room for vagueness!

Secondly, building a model that implements a theory provides a means of testing the internal self-consistency of the theory. A theory that is in any way inconsistent or incomplete will become immediately obvious when trying to implement it as a computer program. The inconsistencies will lead to conflict situations in which the computer program will not be able to function. Such failures point to a need to reassess the situation and to the re-evaluate the theory.

One implication of these two points is that the model can be used to work out unexpected implications of a complex theory. Because the cognitive system operates in a highly complex world, with a multitude of information sources constantly interacting, even a simple process theory can lead to uninterpretable behaviors. Here again, the model provides a tool for teasing apart the nature of these interactions and corroborating or falsifying the theory. One of the earliest applications of artificial neural networks to cognitive processing (McClelland and Rumelhart 1981) was able to demonstrate how constrained interactivity could solve the following puzzle: it is easier to recognize a written letter when it is presented in the middle of a word than in a nonsense string—but how does the reader know that the letter is in a word or a nonsense string before having recognized it?

Perhaps the main contribution made by computational models of cognitive development is to provide an account of the representations that underlie performance on a task that *also* incorporates a mechanism for representational change. One of the greatest unanswered questions of cognitive development is the nature of the transition mechanisms that can account for how one level of performance is transformed into the next level of performance at a later age. This is a difficult question because it involves observing how representations evolve over time, and tracking the interactions between the developing components of a complex cognitive system. Building a model and observing how it evolves over time provides a tangible means of doing this.

Formulating development in computational terms forces the theoretician to be explicit about the transitional mechanisms that underlie information processing. Piaget's own mechanistic theory provides an excellent example of why this is necessary. He described cognitive development in terms of three processes: assimilation, accommodation, and equilibration. Assimilation consisted in adapting or filtering incoming information to make it more compatible with existing knowledge representations. In contrast, accommodation consisted in adapting one's knowledge representations to make them more consistent with novel information. Equilibration was the process by which assimilation and accommodation interacted to cause cognitive development and can therefore be understood as the play-off between these processes of stability and change. Now, while assimilation and accommodation may capture intuitive notions of what is involved in cognitive development, they are too loosely defined to be of any explanatory value. For example, could this theory predict how many errors should be enough to trigger accommodation? Several artificial neural network (or 'connectionist') computational models of cognitive development have sought to address this vagueness by providing computational implementations of assimilation and accommodation (e.g., McClelland 1995; Shultz, Schmidt, Buckingham and Mareschal 1995). In these models, assimilation corresponds to activation flow through the neural network while accommodation corresponds to updating connection weights or the network architecture to reduce output error.

III. Models of Development in Infancy

Infancy is an ideal age range to begin modeling because infant behaviors are not complicated by the presence of language and sophisticated meta-cognitive strategies. Infant abilities are closely tied to their developing sensori-motor skills.

A. Object-Directed Behaviors

Kant identified objects as a fundamental category of cognition. The ability to represent hidden objects liberates infants from the tyranny of direct perception. It is the first step towards representational thought. Piaget (1954) suggested that infants' progress through 6 stages on the way towards reaching an adult level of understanding of object permanence at the age of 2. Many of Piaget's original findings have been replicated. However, changes in methodology, such as relying on where infants look rather than their ability to manually reach for objects, have suggested that infant's understanding of hidden objects is far more precocious (Mareschal 2000). These more recent studies have tended to focus on infant competence at different ages but not on the mechanisms of development from one level of competence to the next. How can the infant's knowledge of the permanence of objects *improve*?

There are relatively few computational models of infant object-directed behaviors. Early models adopted a symbolic stance on the mechanisms that drive behavior and were thus implemented in rule-based production systems (e.g., Luger, Bower, and Wishart 1983; Prazdny 1980). Unfortunately, these were basically competence models that described infant behaviors but did not provide a mechanistic account of development. They proposed different sets of rules to describe behavior at different ages, but did not explain how new rules could be acquired or how one set of rules was transformed into another set of rules. More recent symbolic models have turned to attention-based accounts of object processing in an attempt to explain infant behaviors (Simon 1998). Unfortunately, these models still fail (by and large) to implement any account of *how* development might occur.

One mechanistic learning model has implemented a parallel processing version of Piaget's sensori-motor theory of infant development. Drescher (1991) tried to show how the co-ordination of intra- and inter-modal perceptual motor schemas could lead to a single unified representation of object. Perceptual motor schemas were encoded as "context-action-result" rules and implemented in a parallel processing machine. Learning consisted in using marginal probabilities to fill in context and results slots in appropriate perceptual-motor schemas. Although this system developed an intricate network of intra- and inter-modal schemas that mimicked the infant's sensori-motor integration, it did not develop according to the pattern described by Piaget.

A number of connectionist models have been proposed relying on sub-symbolic representations. In one family of models, a partially recurrent autoencoder network learns to predict the reappearance of a stationary object from behind a moving screen that temporarily hides the object (Munakata, McClelland, Johnson and Siegler 1997). Network performance is measured by taking the difference in response of the nodes coding the location of the hidden object when an object should be revealed, and subtracting it from the response of the node when an object should not be revealed. An increase in this difference is interpreted as increased knowledge of hidden objects. This model demonstrates that the knowledge of objects necessary to retrieve them from behind a screen can be graded and arise incrementally though interactions with an environment.

Mareschal, Plunkett and Harris (1999) described an alternative connectionist model that is more closely tied to the neuropsychological finding that knowledge about an object's identity and its location are processed along separate neural pathways. This model uses a combination of modules to implement dual-route processing. One route learns to process spatial-temporal information while the other route learns to process feature information. Finally, a response module recruits and co-ordinates the representations developed by the other modules as and when required by a response task such as reaching. The specialization of the two routes is initially defined only in terms of different associative learning mechanisms that act on the same input. Empirical studies inspired by this developmental model have subsequently found strong evidence for a dissociation between location information and identity information in the memory of young infants for hidden objects (Mareschal and Johnson 2003).

In contemporary psychological theory, normal cognitive development is increasingly seen within the context of the ways that development can go wrong or operate sub-optimally.

B. Perceptual Categorization

Categorization lies at the heart of cognition because it is the process we relate new individual experiences to our existing knowledge. It is therefore not surprising to find that great deal of effort has been exerted in trying to understand the early roots of category formation. Many infant categorization tasks rely on preferential looking or habituation techniques, based on the finding that infants direct more attention to unfamiliar or unexpected stimuli (Mareschal and Quinn 2001). In a preferential looking experiment, the infant is offered two stimuli to look at: preference for one indicates an ability to distinguish between them. Habituation relies on the fact that infants become board by the repetition of a sequence of identical stimuli. Re-engagement with a new stimulus is evidence that the infant can distinguish it from the previous items. Connectionist autoencoder networks have been used to model the relation between sustained attention and the real-time construction of mental representations in the infant (Mareschal, French and Quinn 2000). The successive cycles of training in the autoencoder reflect an iterative process by which a reliable internal representation of the visual input is established. This approach assumes that infant looking times are positively correlated with the network error. That is, the greater the error, the more novel the stimulus, because it takes more training cycles to reduce the error. The more novel the stimulus, the longer the looking time.

The perceptual categories formed by infants are not always the same as the corresponding adult categories. For example, when 3 to 4-month-old infants are shown a series of cat photographs, they will form a category of CAT that includes novel cats and excludes dogs (as will adults). Thus, after a series of cats, a novel cat will not be interesting to the infant but a novel dog will be. However, when shown a series of dog photographs, the same infants will form a category of DOG that includes novel dogs *but also* includes cats (in contrast to adults). Many aspects of early infant perceptual categorizations (including this asymmetric exclusivity of CAT and DOG categories) are captured by the connectionist autoencoder model. While adults apply top-down schemas when recognizing photographs of cats and dogs, both 3- to 4-

month-olds and the autoencoder networks simply process the bottom-up information in these images. Hence, their internal category representations are yoked to the distributional properties of features in the images. The model validates the idea that categorical representations can self-organize in a neural system as a result of exposure to the familiarization exemplars encountered within the test session itself.

Computer models are invaluable tools for transforming developmental psychology from a descriptive science into a mature explanatory science.

This model can also be used to make new predictions about what kinds of categories infants will form when presented with particular Cat and Dog pictures. French and colleagues (French, Mareschal, Mermillod and Quinn 2004) generated sets of cat and dog pictures that all looked equivalent to adults, but which led to very different categorization behaviors in the infants. Because the infants were attending to low-level features of the image, their responses could be manipulated by changing the distribution of these features. This was not the case with the adults who, driven by their high-level schemas, continued to respond to all images in the same way.

IV. Models of Development in Childhood

Language acquisition marks the end of infancy and the beginning of childhood. Reasoning and conceptual development are the hallmark of cognitive development in childhood. The models in this section all focus on some aspect of reasoning development. We begin by reviewing models that have explicitly tried to implement Piagetian ideas, that is, development that passes through a sequence of stages of increasingly sophisticated reasoning. This is followed by a review of work that breaks away from the Piagetian tradition.

A. Modeling Piagetian Stage Development

There have been several attempts to explain the apparent stage-like growth of competence in children in terms of self-organization in dynamic systems, competition between cognitive growers, and bifurcation theory (e.g., van der Maas and Molenaar 1992; van Geert 1998). However such accounts have tended to rely only on mathematical descriptions that are either not implemented in running computer models or not grounded in measurable information processing components of the cognitive system.

Researchers trying to implement Piagetian notions of development have applied their models to simulating children's performance on key tasks. One such task is known as *conservation*, where children learn that certain properties of objects are preserved through transformations while others are not. Thus, the volume of water is not altered by transferring it between different shaped jugs. Several models have applied themselves to the process of learning these invariant properties under transformation (Klahr and Wallace 1976; Richardson, Forrester et al., 2006; Shultz 1998). Another such task is the *seriation* (or sorting) task. Piaget found that children's ability to order a set of sticks according to length developed through a number of stages. In a first stage, children were unable to sort the sticks. In a second stage, they were able to apply local ordering relations but could not extend the order to the set as a whole. In the third stage, they were able to sort the set of sticks, but only by applying a costly trial and error strategy. Finally, in the fourth stage, children were able to sort the set quickly and efficiently by applying a systematic selection strategy.

Young (1976) approached this task from an information processing perspective. He carried out detailed analyses of the actions children carried out at different ages when sorting blocks. Based on the results of protocol analyses, he developed a rule-based production system that captured children's performance at each stage of development. Progress from one stage to the next was modeled by the (hypothesized) modification of the rules. Although this model provided a good fit to children's behavior at individual stages, the model does not include a working account of how those rules are modified. By contrast, a more recent connectionist model of the seriation task explicitly demonstrates how development could occur (Mareschal and Shultz 1999). In this model, development consists in the gradual tuning of connection weights as the model is exposed to stick-sorting problems of varying complexity. As it learns, the model exhibits a gradual extension of knowledge about small sets to larger sets. It not only captures the stage progression described by Piaget, but it also captures the variability in sorting behaviors observed both within and between different children.

B. Beyond Piaget: The Balance-Scale Task

A recent benchmark of cognitive development is the *balance-scale* task. This was first developed by Inhelder and Piaget in the 1950s and later significantly extended by Robert Siegler. Siegler (1976) explored children's developing abilities to reason about a balance scales (like a scaled-down version of a see-saw). In these problems, children were presented with a symmetric balance scale with 5 equally spaced pegs on either side of the fulcrum. Different numbers of weights were then placed on pegs to the left and the right of the fulcrum and children were asked to predict whether the balance scale would tip to the left, to the right, or remain balanced.

Children show an increasingly sophisticated ability to reason about these problems with increasing age. Seigler (1976) demonstrated that children's strategies at different ages could be characterized by different rules. Rule 1 children, for instance, relied only on a dominant dimension (weight) to predict which side the balance scale would tip. These children would predict that the side with the most weights would be the side that the balance scale would tip. Rule 2 children would apply the same rule as the proceeding children, but had an additional rule stating that if the number of weights was equal on both side, the side with the greatest distance would predict the side on which the balance scale would tip. Rule 3 children behaved like the Rule 2 children with the exception that if the weight and distance cues provided conflicting answers they would guess which side went down. Finally, Rule 4 children had a set of rules that effectively computed the torque on both sides of the balance scale (the number of weights multiplied by the distance from the fulcrum) and chose the side with the greatest torque. Seigler suggested that this knowledge was represented in the form of a growing decision tree. Klahr (1992) pointed out the equivalence of this representational format to rules—a format consistent with production system models of cognitive development.

Both connectionist (McClelland 1995; Richardson, Baughman et al., 2006; Shultz, Mareschal and Schmidt 1994) and decision-tree approaches (Schmidt and Ling 1997) have subsequently been applied to capturing the developmental stages of the balance scale task have been proposed. The connectionist models construe the problem as one of integrating information from two sources. They capture stage development in terms of continuous gradual weight changes and/or recruitment of additional computational resources (internal units) to advance learning. An assumption of these models is that children have greater experience with weight comparisons than distance comparisons. The decision tree model uses the C4.5 tree-inducing algorithm. Children were hypothesized to have an increasing memory capacity and to care increasingly about the detailed correctness of their answers. While this latter model captures the main features of children's performance, the decision trees it developed did not map onto those proposed by Siegler to reflect children's knowledge at different ages. Simulations of children's performance on the balance scale task bring into relief the distinction between *symbolic*, rule-based models and *sub-symbolic*, network-based models, even when it is agreed that the behavior itself can be described as rule following.

C. Modeling of the Development of Reasoning

Many of the successful developmental models described above are connectionist models. Such models process information based on the surface similarity between different exemplars. However, there are cases when children's (and indeed adults') reasoning does not follow surface similarity. Analogical reasoning, for example, requires the child to distance his or herself from the surface similarity between the target and vehicle domains. Solving a problem such as "A cat to a kitten is like a dog to a ?" probes for knowledge of abstract relations like *offspring*. This contrasts with the example of infant categorization, where relationships between cats and dogs depending on the similarity of visual features. For analogical reasoning problems, children move from basing their analogies on surface similarity between the two domains (such as color) to structural similarity (such as the function of an object) between the ages of 6 and 9 years. On the face of it, this reliance on structural similarity is difficult for connectionist systems to capture, given that the models we have described are driven by learned associations between surface features.

Researchers have begun to use 'synthetic brain imaging' as a way to observe the fluid dynamics of damage and recovery in artificial neural networks.

Gentner and colleagues (Gentner et al., 1995) have suggested that adults and children solve analogical problems by comparing mental representations via a structure-mapping process of alignment of conceptual representations. A structurally consistent match conforms to a one-to-one mapping constraint between the domains. For our example, the relation *offspring* derived from the cat-kitten comparison is combined with dog to access the structured representation <dog *has offspring* puppy>. The process is implemented in the Structural Mapping Engine (SME) model. This SME is used to model the relational shift in children's analogical reasoning in terms of increased domain knowledge. As their knowledge of domain relations increases and concepts like *offspring* are acquired, so children's relational representations within a domain become richer and deeper. This increases the likelihood that their comparisons will focus on matching relations rather than surface features. In this view, what develops between 6 and 9 years is only knowledge and not processing.

Recent attempts to explain the origins of analogical reasoning in terms of connectionist neural network properties have de-emphasized the importance of structural alignment in analogical completion. Leech and colleagues (e.g., Leech, Mareschal and Cooper 2003) have suggested that, at least in young children, a form of semantic priming in a large semantic network can best explain analogical completion of the sort described above.

Shrager and Seigler (1998) presented a model of strategy choice, with the intention of providing an account of the range and variability of strategies observed in young children's problem solving. Their model addressed the strategies used by children when adding integers. The strategies are explicitly represented in terms of rules. Strategy choice is probabilistic. The probability of retrieving and executing a strategy depends on the previous association of that strategy with an outcome in conjunction with considerations of cost and efficiency. The strategy pool evolves according to a Darwinian procedure in which in frequently used strategies die off and new strategies enter the pool via random perturbation of existing strategies.

V. Understanding Information Processing in the System

One drawback of simulations of complex systems is that the interactions underlying the model's overt behavior can be opaque to the modeler, and therefore compromise theory development. Models that simulate human behavior but are themselves impenetrable are of no use to the psychologist. In symbolic, rule-based models, the explicit rules that drive behavior offer transparent explanations. However, sub-symbolic models offer a greater challenge and motivate new methods for understanding system performance. For example, researchers have begun to use 'synthetic brain imaging' as a way to observe the fluid dynamics of damage and recovery in artificial neural networks (e.g., Cangelosi and Parisi 2004). Figures 1 and 2 exemplify this approach, using the simulations of Thomas and Karmiloff-Smith (2002) that explored the consequences of early brain damage. Those researchers used a particular aspect of language development, the English past tense, as a test domain. The English past tense has a dual nature, include both regular (talk-talked) and irregular (sing-sang, go-went, hit-hit) past tense forms. Some researchers argue that the cognitive system uses different mechanisms to learn each type of verb (Pinker 1991) and even that there are separate areas in the brain for each mechanism (Tyler, Marslen-Wilson, and Stamatakis 2005). Other researchers argue that sub-symbolic connectionist systems can acquire both regular and irregular mappings in the same network. Thomas and Karmiloff-Smith (2002) demonstrated that in a dual-route network, emergent specialization of function could occur so that across development, one route came to specialize in processing regular verbs while the other processed irregular verbs. Specialization is driven by the greater complexity of irregular mappings and the fact that only one route in this model included the hidden units that are necessary to learn linearly inseparable irregular mappings.

The nature-nurture debate has thus been rephrased into a question of how detailed the initial computational constraints must be in the cognitive system, given the latent structure that we now know is present in the physical and social environment to which infants are exposed.

Figure 1 shows a schematic of the architecture of this dual-route backpropagation network. This is a standard feed-forward connectionist neural network that is trained to associated the stem of a verb with its English past tense. Figure 2 (left column) uses the synthetic brain-imaging technique to demonstrate the involvement of each of the two routes in producing the regular past tense across development. The colors depict how hard each route is driven by the input. The other two columns reveal what happens in the network when it experiences damage to either the left or right route prior to training.

The results highlight the dynamics of development and compensation. Under conditions of normal development, there is partial specialization of function to different components of the system (e.g., specialization in the left or right brain hemispheres by late childhood). However, when one component is damaged early in development by lesioning a large proportion of the connections, the other component is able to compensate to some extent, taking on the overall function. Nevertheless, residual resources in the damaged component may still be recruited where possible.

There is some evidence for this in recovery after early child brain damage (Liégeois et al., 2004). After early unilateral brain damage, children are delayed in their development but usually recover to fall within the normal cognitive range by adolescence. However, there may remain subtle deficits in behavior that depend on the original side of damage, an effect also exhibited by the dual-route

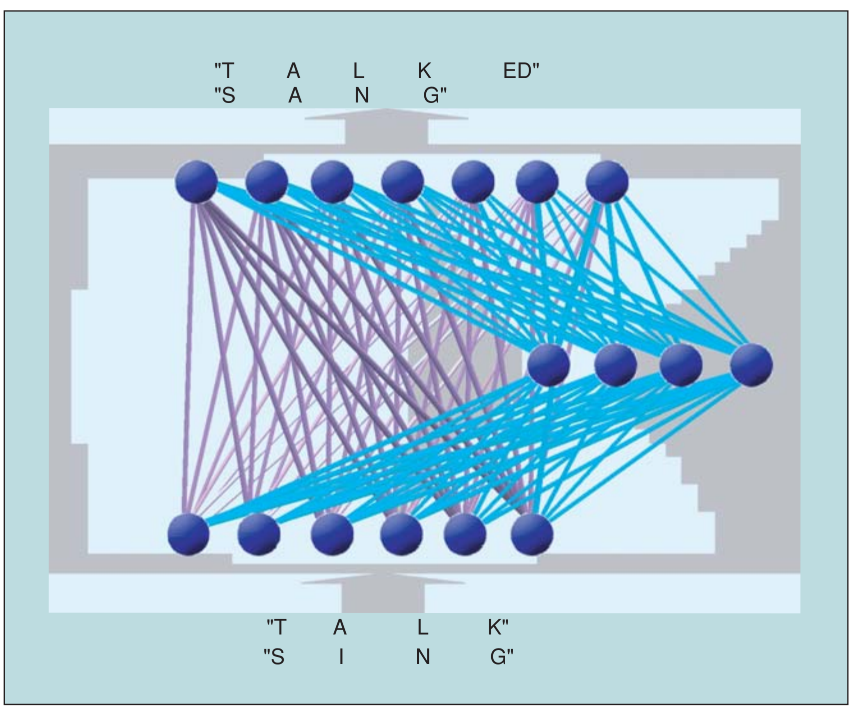

FIGURE 1 Architecture of an associative connectionist network learning the English past tense problem, used to study the dynamics of damage and recovery in developing cognitive systems (Thomas and Karmiloff-Smith 2002).

model. Interestingly, these hemisphere-dependent deficits are more prominent in the development of visuo-spatial processing than in language processing (Bates and Roe 2001; Stiles 2005). The fluid compensation exhibited by children after early focal brain damage and captured in this model is less evident in adults, where there appears to be reduced scope for reorganization. A related avenue of computational modeling work has therefore begun to explore the factors that explain why the plasticity of the cognitive system appears to reduce with age (Thomas and Johnson 2006).

Formal systems offer the ideal framework, and indeed perhaps the only viable method, with which to capture the multiple influences that shape the genesis of the human mind.

VI. Challenges to Current Models of Cognitive Development

We have seen how a shift to more formal, computational approaches to cognitive development has forced psychologists to confront some of the hard theoretical problems, and in particular, the nature of transitional mechanisms. Models differ in their appeal to symbolic or sub-symbolic forms of computation. Symbolic computation offers a better characterization of the abstract reasoning skills observed in older children and adults and of, for example, the complexities of syntax in language. However, models relying on rule-based formats have continued to struggle to elucidate the process by which more complex rules are acquired. The sub-symbolic approach makes development seem more straightforward, since it involves the tuning of continuously valued parameters through experience. Yet its domains of greatest success tend to be those involving the association of features and pattern recognition skills. The reconciliation of these two approaches is a challenge that lies ahead.

The nature of representation is tied up with two other long-standing debates. How much of our cognitive system is genetically specified, and how much is a product of our cultures and environments? Modeling has not settled this issue but offers tools to construe what 'genetically specified' might mean. In the context of capturing behavioral deficits in genetic developmental disorders, for instance, it corresponds to computational constraints that affect the trajectories of learning. The nature-nurture debate has thus been rephrased into a question of *how detailed* the initial computational constraints must be in the cognitive system, given the latent structure that we now know is present in the physical and social environment to which infants are exposed. These days, the focus is on exploring how the process of development could happen given a particular learning system and a particular training set.

Finally, while computer models allow us to formulate questions about what can possibly cause cognitive development (e.g., changes in processing capacity, processing speed, knowledge, and strategy choice) more constraints are required to identify the actual mechanisms involved in children's cognitive development. Recent advances in neuroimaging techniques have allowed us to place greater constraints on how information is processed in the brain. Many of the models above make little use of these constraints and, in the future, such constraints should be incorporated in any functional models of development (Westermann, Sirois, Shultz, and Mareschal 2006). Furthermore, cognitive development does not occur in a social vacuum. Vygotsky (1978) has emphasized the role of social interactions in cognitive development. Society provides a

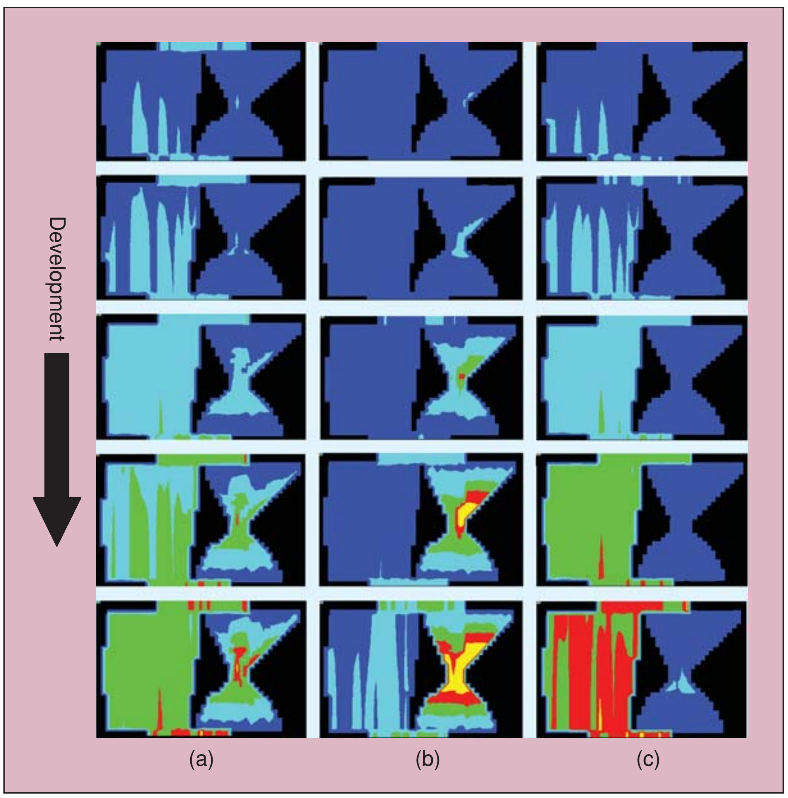

FIGURE 2 The use of 'synthetic brain imaging' to explore the involvement of each route in processing regular verbs across development under normal conditions (column A), and after initial unilateral damage to the left route (column B) or right route (column C). Damage corresponded to lesioning 80% of connections. Colors demonstrate how hard each route is driven by the input (dark blue = least, yellow = most). Formally, values correspond to the product of activations along each weight and the magnitude of the weight, summed for each unit. Excitation and inhibition are therefore treated equivalently; they are also indistinguishable in human brain imaging techniques.

kind of cognitive scaffolding that nurtures and aids the child's cognitive development by actively selecting and filtering the type of problems the child is faced with at any age. Future models will need to consider these constraints to reflect the child's learning environment more accurately. Indeed, significant steps are being made towards developing models that incorporate constraints from multiple levels of description . . . be it at the cellular level the functional brain system level, the cognitive level or the social level of development (Mareschal et al, in press). Formal systems offer the ideal framework, and indeed perhaps the only viable method, with which to capture the multiple influences that shape the genesis of the human mind.

Acknowledgments

This research was supported by European Commission Framework 6 grant 516542-NEST, and by UK MRC grant (G0300188) awarded to MT.

References

[1] E. Bates and K. Roe, "Language development in children with unilateral brain injury," In C.A. Nelson and M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience,* Cambridge, MA: MIT Press, pp. 281–307, 2001.

[2] M.A. Boden, *Artificial Intelligence in Psychology*. Cambridge, MA: MIT: Press, 1988.

[3] M.A. Boden, *Piaget* (2nd edition), Modern Masters, Fontana Press, 1995.

[4] P. Buttery, "A quantitative evaluation of naturalistic models of language acquisition: The efficiency of the triggering learning algorithm compared to a categorical grammar learning," In W. Sakas (Ed.), *Proceedings of the COLING 2004 Workshop "Psychocomputational Models of Human Language Acquisition"* Geneva: COLING, pp. 1–8, 2004.

[5] A. Cangelosi and D. Parisi, "The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging," *Brain and Language*, vol. 89, no. 2, pp. 401–408, 2004.

[6] G.L. Drescher, *Made-up Minds. A Constructivist Approach to Artificial Intelligence,* MIT Press, 1991.

[7] D.J. Field, "What is the goal of sensory coding?" *Neural Computation,* vol. 6, pp. 559–601, 1999a.

[8] D.J. Field, "Wavelets, vision and the statistics of natural scenes." *Philosophical Transactions of the Royal Society of London A,* vol. 357, pp. 2527–2542, 1999b.

[9] J. Fodor, "Fixation of belief and concept acquisition," In: Piatelli-Palmarini, M (ed.) *Language and learning: The debate between Chomsky and Piaget*, Harvard University Press, pp. 143–149, 1980.

[10] R.M. French, D. Mareschal, M. Mermillod, and P. Quinn, "The role of bottom-up processing in perceptual categorization by 3- to 4-month old infants: Simulations and data," *Journal of Experimental Psychology: General,* vol. 133, 382–397, 2004.

[11] D. Gentner, M.J. Rattermann, A.B. Markman, and L. Kotovsky, "Two forces in the development of relational similarity," In T.J. Simon and G.S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling*, Hillsdale, NJ: LEA. pp. 263–313, 1995.

[12] A. Karmiloff-Smith, "Development itself is the key to understanding developmental disorders," *Trends in Cognitive Sciences*, vol. 2, 389–398, 1998.

[13] D. Klahr and J.G. Wallace, *Cognitive development: An information processing view*, Hillsdale, NJ: Erlbaum, 1976.

[14] D. Klahr, "Information processing approaches to cognitive development," In Bornstein MH and Lamb ME (Eds.), *Developmental psychology: An advanced textbook* (3rd ed.), Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 273–335, 1992.

[15] R. Leech, D. Mareschal, and R. Cooper, "A connectionist account of analogical development," In R. Altman and D. Kirsh (eds.) *Proceedings of the twenty-fifth annual conference of the Cognitive Science Society,* London: LEA, pp. 710–715, 2003.

[16] F. Liégeois, A. Connelly, J. Helen Cross, S.G. Boyd, D.G. Gadian, F. Vargha-Khadem, and T. Baldeweg, "Language reorganization in children with early-onset lesions of the left hemisphere: An fMRI study," *Brain*, vol. 127, no. 6, pp. 1229–1236, 2004.

[17] G.F. Luger, T.G.R. Bower, and J.G. Wishart, "A model of the development of the early object concept," *Perception*, vol. 12, pp. 21–34, 1983.

[18] D. Mareschal, "Infant object knowledge: Current trends and controversies" *Trends in Cognitive Science*, vol. 4, pp. 408–416, 2000.

[19] D. Mareschal, R.M. French, P. Quinn, "A connectionist account of asymmetric category learning in infancy," *Developmental Psychology*, vol. 36, pp. 635–645, 2000.

[20] D. Mareschal and M.H. Johnson, "The "What" and "Where" of infant object representations," *Cognition*, vol. 88, pp. 259–276, 2003.

[21] D. Mareschal, M.H. Johnson, S. Sirois, M. Spratling, M. Thomas, and G. Westermann, *Neuroconstuctivism Vol. 1: How the brain constructs cognition*, Oxford UK: OUP, in press, 2006.

[22] D. Mareschal, K. Plunkett, and P. Harris, "A computational and neuropsychological account of object-oriented behaviors in infancy," *Developmental Science*, vol. 2, pp. 306–317, 1999.

[23] D. Mareschal and P.C. Quinn, "Categorization in infancy," *Trends in Cognitive Science,* vol. 5, pp. 443–450, 2001.

[24] D. Mareschal and T.R. Shultz, "Children's seriation: A connectionist approach," *Connection Science,* vol. 11, pp. 153–188, 1999.

[25] D. Mareschal and M.S.C. Thomas, "What are developmental psychologists looking for in a model?" *IEEE Transactions on Evolutionary Computation,* Special Issue on Autonomous Mental Development, in press.

[26] J.L. McClelland, A connectionist perspective on knowledge and development, In T. Simon and G. Halford (Eds.), *Developing Cognitive Competence: New Approaches to Process Modeling*, Hillsdale, NJ: Erlbaum, pp. 157–204, 1995.

[27] J.L. McClelland and D.E. Rumelhart, "An interactive-activation model of context effects in letter perception: Part 1, an account of basic findings," *Psychological Review*, vol. 88, pp. 375–407, 1981.

[28] Y. Munakata, J.L. McClelland, M.H. Johnson, and R.S. Siegler, "Rethinking infant knowledge: Towards an adaptive process account of successes and failures in object permanence tasks," *Psychological Review*, vol. 104, pp. 686–713, 1997.

[29] J. Piaget, *Psychology and epistemology: Towards a theory of knowledge*. Middlesex, UK: Penguin Books, 1971.

[30] J. Piaget, *The Construction of Reality in the Child*. New York: Basic Books, 1954.

[31] S. Pinker, "Rules of language," *Science*, vol. 253, pp. 530–535, 1991.

[32] D.C. Plaut, J.L. McClelland, M.S. Seidenberg, and K.E. Patterson, "Understanding normal and impaired word reading: Computational principles in quasi-regular domains," *Psychological Review*, vol. 103, pp. 56–115, 1996.

[33] S. Prazdny, "A computational study of a period of infant object-concept development," *Perception,* vol. 9, pp. 125–150, 1980.

[34] F.M. Richardson, F.D. Baughman, N.A. Forrester, and M.S.C. Thomas, "Computational modeling of variability in the balance-scale task," *Proceedings of the 7th International Conference of Cognitive Modeling,* Trieste, Italy: Edizioni Goliardiche, pp. 256–261, 2006.

[35] F.M. Richardson, N.A. Forrester, F.D. Baughman, and M.S.C. Thomas, "Computational modeling of variability in the conservation task," In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Vancouver, BC, Canada, July 26–29, 2006.

[36] B. Rogoff, *Apprenticeship in thinking*. Oxford, UK: Oxford University Press, 1990.

[37] W.C. Schmidt and C.X. Ling, "A decision-tree model of balance scale development," *Machine Learning*, vol. 18, pp. 1–30, 1996.

[38] J. Shrager and R.S. Siegler, "SCADS: A model of children's strategy choices and strategy discoveries," *Psychological Science*, vol. 9, pp. 405–410, 1998.

[39] T.R. Shultz, "A computational analysis of conservation," *Developmental Science*, vol. 1, pp. 103–126, 1998.

[40] T.R. Shultz, D. Mareschal, and W.C. Schmidt, "Modeling cognitive development on balance scale phenomena," *Machine Learning*, vol. 16, pp. 59–88, 1994.

[41] T.R. Shultz, W.C. Schmidt, D. Buckingham, and D. Mareschal, Modeling cognitive development with a generative connectionist algorithm, In T. Simon and G. Halford (Eds.), *Developing Cognitive Competence: New Approaches to Process Modeling*, Hillsdale, NJ: Erlbaum, pp. 347–362, 1995.

[42] R.S. Siegler, "Three aspects of cognitive development," *Cognitive Psychology,* vol. 8, pp. 481–520, 1976.

[43] T.J. Simon, "Computational evidence for the foundations of numerical competence," *Developmental Science*, vol. 1, pp. 71–78, 1998.

[44] J. Stiles, J. Reilly, B. Paul, and P. Moses, "Cognitive development following early brain injury: Evidence for neural adaptation," *Trends in Cognitive Sciences*, vol. 19, no. 3, pp. 136–143. 2005.

[45] M.S.C. Thomas and M.H. Johnson, "The computational modeling of sensitive periods," *Developmental Psychobiology,* vol. 48, pp. 337–344, 2006.

[46] M.S.C. Thomas and A. Karmiloff-Smith, "Are developmental disorders like cases of adult brain damage? Implications from connectionist modeling," *Behavioral and Brain Sciences*, vol. 25, no. 6, pp. 727–780, 2002.

[47] M.S.C. Thomas and A. Karmiloff-Smith, "Connectionist models of development, developmental disorders and individual differences," In R.J. Sternberg, J. Lautrey, and T. Lubart (Eds.), *Models of Intelligence: International Perspectives*, American Psychological Association. pp. 133–150, 2003.

[48] L.K. Tyler, W.D. Marslen-Wilson, and E.A. Stamatakis, "Differentiating lexical form, meaning and structure in the neural language system," *Proceedings of the National Academy of Sciences*, vol. 102, no. 23, pp. 8375–8380, 2005.

[49] H. Van der Maas and P. Molenaar, "A catastrophe-theoretical approach to cognitive development," *Psychological Review*, vol. 99, pp. 395–417, 1992.

[50] P. van Geert, "A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond," *Psychological Review*, vol. 105, pp. 634–677, 1998.

[51] L.S. Vygotsky, *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press, 1978.

[52] G. Westermann, S. Sirois, T.R. Shultz and D. Mareschal, "Modeling developmental cognitive neuroscience," *Trends in Cognitive Sciences,* vol. 10, pp. 227–232, 2006.

[53] R. Young, *Seriation by children: An artificial intelligence analysis of a Piagetian task*. Basel: Birkhauser 1976.