Computational Modeling in Developmental Psychology

Denis Mareschal and Michael S. C. Thomas

Abstract—**This manuscript surveys computational modeling efforts by researchers in developmental psychology. Developmental psychology is ready to blossom into a modern science that focuses on causal mechanistic explanations of development rather than just describing and classifying behaviors. Computational modeling is the key to this process. However, to be effective, models must not only mimic observed data. They must also be transparent, grounded, and plausible to be accepted by the developmental psychology community. Connectionist model provide one such example. Many developmental features of typical and atypical perception, cognition, and language have been modeled using connectionist methods. Successful models are closely tied to the details of existing empirical studies and make concrete testable predictions. The success of such a project relies on the close collaboration of computational scientists with empirical psychologists.**

Index Terms—**Children, cognitive modeling, connectionism, developmental psychology, infancy.**

I. INTRODUCTION

THE computational study of *development* can have a range of purposes. One purpose is to ask whether computers can develop mental abilities from experience; that is, are the computational properties implemented in the computer sufficient to acquire the target behavior given some series of learning events. In fact, almost 50 years ago, Simon and Newell had already argued that computational systems could acquire and express intelligent behavior. They wrote that "*Intuition, insight, and learning are no longer exclusive possessions of humans: any large high-speed computer can be programmed to exhibit them also. (…) The simplest way [we] can summarize the situation is to say that there are now in the world machines that can think, that learn, and that create*" [92, pp. 6–8]. A second purpose is to use the computer as a tool for understanding the child. In this case, we ask whether a computer exhibiting intelligent behavior is acquiring the ability in the same way as the child. In this capacity, computers can help elucidate the principles by which developmental change occurs in children.

Both approaches have their value and, in fact, can benefit immensely from the cross-fertilization of ideas arising from each domain. Machine-learning approaches to problem solving can generate new hypotheses for how humans achieve the same feats. The study of human development can offer novel

The authors are with the School of Psychology, Birkbeck University of London, London WC1E 7HX, U.K. (e-mail: d.mareschal@bbk.ac.uk; m.thomas@bbk.ac.uk).

engineering solutions to the acquisition of complex abilities. Given this potential for cross-fertilization, there is value in bringing both kinds of effort together in the same community. With this goal in mind, in this paper, we will focus on the latter of the two roles of computational models of development. We do this because it reflects the approach that we use as research psychologists.

For much of its history, developmental psychology has charted what infants and children could and could not do at any given age. The origins of this approach lie deep within the Piagetian tradition of identifying and attempting to draw inferences based on the interesting errors in reasoning that children sometimes make [23], [63], [64]. However, as the 19th century physical chemist Lord Rutherford once said, "*all science is either physics or stamp collecting*" (cited in [4]). A mature modern science is one that strives to understand the causal mechanisms that explain the relationship between a series of observed phenomena. Thus, *scientific* developmental psychology must elucidate the causal mechanisms that underpin the behaviors we see; not just the moment-by-moment observation of children, but also the mechanisms that explain the gradual developmental change that is observed as infants and children grow older [50].

Most modern sciences have progressed from a descriptive to a causal-mechanistic stage. This has been true in the physical sciences, the biological sciences as well as economics and other social sciences. Developmental psychology is now mature enough (in terms of having enough data) to begin the transition into a causal mechanistic science. The principal historical obstacle to this transition has been the lack of an appropriate tool to think about information-processing mechanisms and the processes of developmental change. The introduction into the study of cognitive psychology of computational modeling methods in the mid 1970s and 1980s (see [7]), swiftly followed by the arrival of neural network connectionist modeling [81], has provided just such a tool.

Implemented computational models bring many benefits to the study of psychology [42]. Their key contribution is that they force the researcher to be explicit about the information-processing mechanisms that underlie performance on a task. As such, they test the internal consistency of any proposed information-processing theory and allow the researcher to explore ranges of behaviors that may be impossible or unethical to explore empirically. They can also be used to predict performance under extreme limiting conditions, and to explore the complex interactions of the multiple variables that may impact on performance.

Computer models complement experimental data gathering by placing constraints on the direction of future empirical investigations. First, developing a computer model forces the user to

specify precisely what is meant by the terms in his or her underlying theory. Terms such as representations, symbols, and variables must have an exact definition to be implemented within a computer model. The degree of precision required to construct a working computer model avoids the possibility of arguments arising from the misunderstanding of imprecise verbal theories.

Second, 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 one tries to implement it as a computer programmed: the inconsistencies will lead to conflict situations in which the computer programmed will not be able to function. Such failures point to a need to reassess the implementation or reevaluate the theory.

An implication of these two points is that the model can be used to work out unexpected implications of a complex theory. Because the world is highly complex with a multitude of information sources that constantly interact, even a simple process theory can lead to unforeseen behaviors. Here again, the model provides a tool for teasing apart the nature of these interactions and corroborating or falsifying the theory.

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, while *also* incorporating a mechanism for representational change (see [45], for an extensive discussion of this issue). 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, often of a more complex or abstract nature, at a later age. How can learning produce increases in the sophistication of reasoning? This is a difficult question because it involves observing how representations evolve over time and tracking the intricate interactions between the developing components of a complex cognitive system and its subjective environment. Building a model and observing how it evolves over time provides a tangible means of achieving this end. Indeed, models that do not address transitions but only explain behavior at discrete ages are *not* models of development, even if the relevant data that they explain involves children.

There are of course a number of pitfalls that confront any modeler. The first is making the right choice of which variables to exclude and which variables to focus on in any particular model. A model is, by definition, an approximation of the real system designed quite intentionally to simplify the subject of study. What the right approximations are depends on what the question is, who the target audience is, and what level of understanding has already been reached regarding this field. In his extremely lucid chapter "*Artificial Intelligence Meets Natural Stupidity*," McDermot [52] identifies several further pitfalls of using AI systems as models of human cognition. The first is to attach meaningful labels to components of a model (e.g., "understanding module," "reasoning module," "attachment module," "semantics module," "syntax module"), and then to assume that the system provides some kind of explanation of behavior simply by virtue of using psychological terminology. Such an explanation is illusionary. To show this, McDermot suggests relabeling the components with more abstract labels

(e.g., "Z3XL module" and "A23BP module") to see if the model still provides an explanatory framework for the target behavior.

The second (and unfortunately all too common) pitfall is to describe and draw inferences from unimplemented versions of the model. While it may seem obvious how a given model will perform, until the simulations are actually run, the model has no more value than a standard verbal psychological theory. If the outcome of simulations could always be anticipated in advance, there would be no virtue in modeling. In fact, some of the most creative scientific discoveries have come from unexpected results to what seemed to be blindingly obvious experiments.

For developmental psychologists, a model is fundamentally a tool for helping us to reason about the processes that underlie a given natural phenomenon. To be of value to the developmental community, a number of constraints must be satisfied. The first is *transparency*. A model must be understandable to those who are going to use it in their everyday research activities. This does not mean that its dynamic properties need to be immediately graspable. However, the processes and mechanisms that underlie the system's behavior, their mathematical embodiment and their computational implementation must all be clear.[1] If the end user cannot understand the basic processes underlying the developmental model, then it is of little value, even if it mimics completely the unfolding behaviors observed in a child.

Second, the model must be *grounded*. It must make substantial contact with the rich data already available in all areas of cognitive development. A real danger of interdisciplinary research (such as the computational modeling of cognitive development) is that expertise in one side of a multifaceted project is underrepresented. Researchers then rely on secondary sources to guide their modeling efforts. These secondary sources are often either out of date, of limited scope, or simple approximations of what the real phenomena look like. Consequently, experts in the field do not view the model as making any real theoretical contribution. This is a problem that we have encountered in all aspects of computational modeling. Examples range from psychologists alluding to mathematical concepts such as "Strange Attractors" to explain development without any real understanding of what a Strange Attractor is and what its formal characteristics are [78], [80]. Conversely, computational scientists may develop models of psychological phenomena such as "Dyslexia" or "Autism" without a full understanding of the controversy that surrounds the identification and diagnosis of these disorders and the richness of the range of behaviors that exists in individuals labeled as autistic or dyslexic (e.g., [27]).

Third, the model must be *plausible*. The mechanisms and processes that it proposes must be consistent with those known or believed to exist in other related domains. Putting aside the issue of what the appropriate level of description is for a particular phenomenon (i.e., is it best described at the cognitive level or the neural level of processing?), there is a temptation as engineers and computer scientists to posit mechanisms that will work, but that have no cross-reference to other models in similar domains.

---

[1]Such a proposition is not unique to models of psychological processes. For example, models of highly complex systems such as weather and climate systems are based on fairly straightforward assumptions about physics, and yet these "simple" models help explain otherwise unpredictable climate and weather conditions.

As a result, the model is theoretically and empirically isolated and it becomes difficult to see how the model could generalize to any other domain. In terms of levels of description, while empirical phenomena can be independently studied at different levels, the levels are not themselves independent. Thus, a theory at the cognitive level cannot include assumptions or mechanisms that cannot be delivered by processes at the neural level; assumptions about innate starting points for cognitive development cannot be inconsistent with what is known about genetic constraints on brain development [45].

Computational models of cognitive development typically permit researchers to explore how some target behavior can be acquired via exposure to a set of experiences. From a practical point of view, this exercise can be construed as just another kind of machine-learning problem, and is therefore subject to the same kind of design choices. These include choosing the type of training experience, the target function to be learned, a representation for the target function, and an algorithm for learning this function from the training examples [55]. The particular computational formulism (be it an artificial neural network, a production system, or a decision tree) along with the input and output representational formats and free internal parameters, determine the representational states or space of hypotheses that can be adopted by the system following its exposure to the training set. From this viewpoint, learning involves searching the hypothesis space to find the hypothesis that best fits the training examples and the prior constraints or knowledge. Each type of system also includes an inductive bias with which it extends knowledge acquired from training examples to classify novel inputs. For instance, the inductive bias for artificial neural networks is that implemented functions are smooth; for decision tree algorithms (such as ID3), the inductive bias is that high information-gain attributes should be positioned close to the root and shorter trees should be preferred over larger trees [55].

In a *cognitive* computational model, psychological empirical data are used wherever possible to constrain choices about representations and training sets. Psychological approaches with a nativist leaning will produce models with additional constraints on their architectures, activation dynamics, input/output representations, or learning algorithms, so that fewer internal hypothesis states are available. Training will serve 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, [8]). Models of a more empiricist bent will have fewer constraints on the hypothesis space, so that the information in the training examples plays a stronger role in determining which hypothesis is selected. For example, some theories of the development of the visual system argue that when one exposures a fairly general self-organizing learning system[2] 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 [21], [22].

Generalization to novel inputs is often of primary interest in computational models of cognitive development, since children acquire flexible behavioral patterns in response to given domains that typically do not require the child to have experienced every possible configuration of that domain. It turns out that generalization tests are often more informative of the solution that a system has reached than its performance on the training data, as many different hypothesis states are consistent with a given training set but these frequently diverge in how they treat novel situations. This point will become relevant later, when we consider developmentally disordered systems, which may appear to have acquired the same cognitive processes with reference to a narrow set of behaviors (equivalent to the training set), but exposure to novel situations reveals that normal-looking behavior has been achieved through atypical underlying representations.

In the rest of this paper, we will illustrate these general principles by focusing on one computational modeling approach that has been extremely successful at producing psychologically relevant computational models of learning and development in infants and children. These include models of typical but also atypical learning and development, where key boundary conditions and resource limitations are found to lie at the heart of the atypical behaviors observed in children.

II. Connectionist Models of Development

Many different architectures have been proposed as psychological process models of development. These include productions system models (e.g., [39], [60], and [114]); decision tree models (e.g., [86]), neural network and connectionist models (e.g., [19], [48], [57], and [89]), hybrid models (e.g., [17]), and temporal-difference learning models (e.g., [105]), to name but a few. It is well beyond the scope of this paper to review all of these approaches. Instead, we will focus on one particular approach that has been particularly successful at modeling development, and consequently, well received by the developmental community. Here, we are referring to connectionist or neural network models of cognitive development.

Connectionist networks are computer models loosely based on the principles of neural information processing [19], [54], [81]. In most cases, they are not intended to be neural models, but rather cognitive information processing models that embody general processing principles such as inhibition and excitation within a distributed, parallel processing system. They attempt to strike the balance between importing some of the key ideas from the neurosciences, while maintaining sufficiently discrete and definable components to allow questions about behavior to be formulated in terms of a high-level cognitive computational framework. However, in some cases, the link with neural information processing is made more explicit (see [111]).

From a developmental perspective, connectionist networks are ideal for modeling because they develop their own internal representations as a result of interacting with an environment [71]. However, these networks are not simply *tabula rasa* empiricist learning machines. The representations that they develop can be strongly determined by initial constraints (or

---

[2]A self-organizing learning system is one that when exposed to an input set, develops a set of representations reflecting the latent similarity structure in that input. It is driven only by its own learning algorithm, without an external error signal. By contrast, in a supervised associative system, the training set comprises input–output pairs and the system must learn to associate the appropriate output pattern with each input.

boundary conditions). These constraints can take the form of different associative learning mechanisms attuned to specific information in the environment (e.g., temporal correlation or spatial correlation), or they can take the form of architectural constraints that guide the flow of information in the system. Although connectionist modeling has its roots in associationist learning paradigms, it has inherited the Hebbian rather than the Hullian tradition [32]. That is, what goes on *inside* the box (inside the network) is as important in determining the overall behavior of the networks as the correlation between the inputs (stimuli) and the outputs (responses).

Connectionist networks are made up of simple processing units (idealized neurons) interconnected via weighted communication lines [33]. As activation flows through the network, it is transformed by the set of connection weights between successive layers in network. Thus, learning (i.e., adapting one's behavior) is accomplished by tuning the connection weights until some stable state of behavior is obtained.[3]

In all cases, these networks adapt in such a way as to internalize structures in the environment. Through adaptation, the connection weights come to encode regularities about the network's environment that are relevant to a task the network must solve. Networks are very sensitive to the distribution statistics of relevant features in their environment. A feedforward network with a single layer of hidden units can approximate arbitrarily well any bounded continuous output response function, given enough hidden units (e.g., [14]). Further details of the similarities between connectionist network learning and statistical learning procedures can be found elsewhere (e.g., [30], [31], and [33]).

Psychologists think of knowledge occurring at two levels in connectionist neural networks (see [57]). On the one hand, there is knowledge stored in the connection strengths, an accumulation of learning events. The connection strengths determine the pattern of activation produced by an input and/or by existing activation propagating inside the system. On the other hand, there is knowledge corresponding to the activations themselves. When activation is processed through the connections, it gives rise to maintained activity, which serves as both a representation of current inputs and a working memory for recently presented information.

Note that many connectionist networks are very simple. They may contain some 100 units or so. This is not to suggest that the part of the brain solving the corresponding task only has 100 neurons. Remember that such models are frequently not intended as neural models, but rather as information-processing models of behavior. The models constitute examples of how systems with similar computational properties to the brain can give rise to a set of observed behaviors. Sometimes, individual units are taken to represent pools of neurons or cell assemblies. According to this interpretation, the activation level of the units corresponds to the proportion of neurons firing in the pool (e.g., [9]). However, to preserve the cognitive interpretation of the model, activation of a unit in a network corresponds to a conceptual state with reference to the domain being modeled, rather than the spike train recorded from a single neuron

somewhere in the brain. This is because neural codes typically employ distributed representations rather than localist representations. In distributed codes, the cognitively interpretable unit is represented by a pattern of activity, not by the activity of a single unit. Since the representations in a connectionist model of a cognitive process may capture similarity relations among patterns used in the brain and among concepts entertained by the mind without the units representing the concepts themselves, there is a sense in which the models exist at a level between the cognitive and the neural level.[4]

Usually, the aim of these models is to capture normative development, that is, the development of the normal or average child. The successful model should pass through the observed stages of development, simulate error types and quantities at each stage, and achieve the adult level of performance. Data may include both accuracy and response times. However, normative development profiles mask a great deal of variability. Some of this occurs within the individual, whereby advances in the sophistication of performance may initially be inconsistent, with backtracking and sensitivity to context, before consolidation. Variability can also be found between individuals of the same age. This is often referred to as *individual differences*. Variability may exceed the "normal" range (statistically defined as standard scores varying between 85 and 115), corresponding to learning disabilities and giftedness, respectively. At the lower end of the normal range, variability can also be observed stemming from diagnosed developmental disorders that often have a genetic origin, such as autism, Down syndrome, Williams syndrome, and Attention Deficit Hyperactivity Disorder. Some learning disabilities appear to have relatively restricted scope with other areas of cognitive development closer to the normal range, such as Specific Language Impairment (SLI), dyslexia, dyspraxia, and dyscalculia.

Implemented computational models provide a framework in which issues of variability can be explored [95], [102]. Once a model of normal development has been established, it is possible to assess the consequences of different types of impairment, thereby generating candidate explanations for the variability. As discussed above, constructing a connectionist model involves making a set of decisions about the input and output representations corresponding to the target cognitive domain, the regime of training experiences, the specific architecture and learning algorithm, and a range of free parameters. Within the framework of the model, these are the constraints that act on or shape the normal developmental trajectory. The normative model employs some constraints that are determined by psychological data (e.g., what representations of word sounds, written words, and meanings are psychologically plausible), while other constraints (e.g., the range of possible parameter values) are set to simulate the empirical data.

It is important to be clear about the role of the free parameters in these models. The free parameters are not adjusted simply to fit models to the experimental data they are supposed to explain. Rather, the free parameters constrain the trajectory of learning. Changes in behavior are driven by iterative, *experience-dependent* changes to a separate set of parameters: the connection strengths (and in some cases, the network architecture). The free

---

[3]Some connectionist modelers also change in the network topology to adapt as part of the learning process (see [89] for a full discussion of this approach).

[4]We are grateful to an anonymous reviewer for this suggestion.

parameters influence the way that experience alters this separate set of parameters. Thus, fitting the training data is not generally a criterion for success in these models, while fitting the behavioral data from the combination of the architecture and the training process generally is. To be informative, simulations must capture both the developmental trajectory and in the end state performance of the system, and that end state performance must normally extend to novel situations.

Of course, the availability of free parameters that can accommodate individual differences might suggest that it is easy to simulate any set of psychological empirical data, especially given that connectionist networks can approximate almost any arbitrary function with enough hidden units. There are a number of responses to this critique. First, the fact that there exists a set of parameters that will allow the network to approximate the responses of children on a task does not imply that this solution is learnable by the network. For example, the system may get stuck in a *local minimum* during training, a suboptimal, partial solution to the problem from which any immediate change causes a worsening in performance before a general solution can be found. Second, even if a response function is learnable, not all ways of learning the problem will produce developmental profiles shown by the children: the power of connectionist models is that their developmental profiles have often been found to track those observed in children. Third, the input–output representations and training regime often turn out to be the strongest constraints in many models. In this case, the free parameters do not carry much of the explanatory weight. Fourth, in practice, a lot of productive modeling work has been carried out despite the availability of free parameters. The import assumptions in models have mostly turned out to be theory-driven rather than relying on the setting of unconstrained parameters.

Given a normative model, we can ask whether there are suboptimal settings of the free parameters that can explain the types of behavioral profiles we see in developmental disorders. Is there a parameter variation that could explain individual differences, such as the level of computational resources? Is there some property of representations of, say, spoken words or written words that might make it hard to learn the mapping between the two, and so capture features of developmental dyslexia? Implemented models of developmental disorders and individual differences offer the same drive to clarify previously vague verbal terms. In this case, they are terms like "processing speed" (differences in which are taken by some to explain the general factor of intelligence in psychometric testing), "poor representations," and "delay" (both offered as explanations of developmental disorders).

From a formal learning perspective, alterations to the model's constraints may produce a number of effects. They may change the nature of the hypothesis space that can be reached; they may change the nature of the search of an existing hypothesis space; they may change the inductive bias which the system uses to generalize to novel examples; or they may change the set of training examples, either in the system's autonomous sampling of the environment or when the environment is itself impoverished.

The utility of computational models for simulating individual *variability* can best be illustrated by a brief example. One ongoing debate in the field of development disorders is their relation to acquired disorders following brain damage [38]. Is a child with the developmental disorder of SLI similar in any way to the adult with acquired aphasia? Modeling was able to generate insights into this question by investigating the consequences of damaging a learning system in its initial state (analogous to a developmental disorder) compared with damaging a system in its trained state (analogous to an adult acquired deficit) [100]. The results demonstrated that some types of damage hurt the system much more in the "adult" state (e.g., severing network connections), while others hurt the system much more in the "infant" state (e.g., adding noise to processing). The adult system can tolerate noise because it already possesses an accurate representation of the knowledge, but loss of network structure leads to a decrement in performance since connections contain established knowledge. By contrast, the infant system can tolerate loss of connections because it can reorganize remaining resources to acquire the knowledge, but is impaired by noisy processing since this blurs the knowledge that the system has to learn. Empirical evidence supports the importance of a good representation of the input during language acquisition. When McDonald [53] analyzed the conditions for successful and unsuccessful language acquisition across a range of populations (including early and late first language learners, early and late second language learners, individuals with Down syndrome, Williams syndrome, and SLI), the results indicated that good representations of speech sounds were key in predicting the successful acquisition of a language, including its syntax.

Computational models of variability can serve a particularly important role where this variability is in part of genetic origin. The developmental pathway between gene expression and children's behavior is extremely remote, since genes only code for proteins (or control the action of other genes) at a cellular level, while behavior is a property of the whole organism potentially occurring many years after the relevant gene expression exerted its role on development. To establish how variability on the genome produces variability in behavior, we must understand how differences in gene expression alter the low-level properties of the brain as it develops, how these differences lead to atypical neurocomputational properties, and how these properties deflect the subsequent developmental trajectory a child follows through engaged interaction with their subjective physical and social environment [38], [45]. Although one might read in a newspaper that " a gene for dyslexia has been found," the causal pathway between the two will only be complete when a theory of how gene expression modulates functional brain development is reconciled with a developmental computational model of learning to read which, under suboptimal computational conditions, struggles to pronounce the word YACHT or the made-up word PLONCH after ample amounts of training.

III. FOUR CONNECTIONIST MODELS OF PSYCHOLOGICAL DEVELOPMENT

A broad range of neural network methods fall within the general framework of Connectionism. These include feedforward and recurrent associative networks (e.g., [19]), autoassociator networks (e.g., [93]), "generative" networks that grow their own architectures as part of the developmental process (e.g., [89]),

adaptive resonance networks (e.g., [74]), and multiagent approaches (e.g., [85]) among others. Interested readers can find out more about these and other approaches that are discussed in more detail elsewhere [50].

Connectionist modeling tools have been used to investigate a wide variety of different domains within development, spanning all the way from early nonverbal infant behaviors to the high-level reasoning of older children. These include infant-object interactions [46], [56], [58], problem solving on the balance-scale task [51], [76], [91], classic Piagetian cognitive stage tasks [49], [88], numerical reasoning [1], and set discrimination learning [94].

A large body of research has also focused on explaining the processes of language development (see [11] and [12], for reviews). Models have employed self-organizing networks to explore processes of early phonological development in learning to categories speech sounds [59], [110]. Recurrent architectures have been employed in models of segmenting the speech stream into discrete words [10] and extracting syntactic structure from sequences of words (e.g., [18]). Generative networks have been employed to model the acquisition of personal pronouns [61]. Feedforward networks and attractor networks have been used in models of the development of vocabulary [26], [72], and of metaphor comprehension [103], as well as in two of the most heavily researched areas of language acquisition, the development of inflectional morphology, including past tense and pluralization (e.g., [24], [28], MacWhinney & Leinbach, 1991; [67]–[70], [82], [101]), and the development of reading (e.g., [29], [66], [87]).

In the following sections, we focus on four examples of connectionist models taken from our own work that illustrate the kind of modeling that is well received by developmental psychologists. Our intention is to provide the readers with a real flavor for the objectives of cognitive computational modeling, as well as the obstacles encountered and their proposed solutions. We begin with two models of early category learning in infancy. One focuses on the infants' behavior as it unfolds within a test session, while a second explores changes in how infants categorize across development. We then describe two models of atypical development. The first examines inflectional morphology (an aspect of language acquisition) in SLI, while the second discusses the development of atypical gaze behavior in infants with autism and Williams syndrome.

A. Perceptual Category Learning in Early Infancy

Categories and concepts facilitate learning and reasoning by partitioning the world into manageable units. Even 3- to 4-month-olds can categorize a range of real-world images. Research by Quinn and Eimas demonstrates that these infants can categorize photographs of cats, dogs, horses, birds, tables, and chairs [47]. However, the perceptual categories do not always have the same characteristics as might be expected from the adult concepts. In particular, the extension and exclusivity of infant categories (i.e., the range of exemplars accepted or rejected as members of the category) may differ from that of adult categories.

Quinn *et al.* [73] used a familiarization/novelty-preference technique to determine if the perceptual categories of familiar animals (e.g., cats and dogs) acquired by young infants would

exclude perceptually similar exemplars from contrasting basic-level categories. They found that when 3- to 4-month-olds are familiarized with six pairs of cat photographs presented sequentially (12 photographs), the infants will subsequently prefer to look at a novel dog photograph rather than a novel cat photograph. Because infants prefer to look at unfamiliar stimuli [20], this was interpreted as showing that the infants had developed a category of Cat that included novel cats (hence, less looking at the cat photograph) but excluded novel dogs (hence, more looking at the dog photograph). However, if the infants are initially familiarized with six pairs of dog photographs sequentially (12 photographs), they will show no subsequent preference for looking at either a novel dog or a novel cat. Furthermore, control conditions revealed that: 1) the infants would prefer to look at a novel test bird after initial familiarization with either dogs or cats; 2) there is no *a priori* preference for dogs over cats; and (3) the infants are able to discriminate within the Cat and Dog categories. Taken together, these findings led Quinn *et al.* to suggest that the 3- to 4-month-olds had formed a perceptual category of Dog that included novel dogs but *also* included novel cats.

Mareschal *et al.* [44] constructed a simple connectionist model to try to explain this behavior in terms of simple associative memory loading multiple exemplars. The model consisted of a standard 10–8–10 feedforward autoencoder network trained using the backpropagation learning algorithm. The data for training the networks were obtained from measurements of the original Cat and Dog pictures used by Eimas and Quinn. There were 18 dogs and 18 cats classified according to the following ten traits: head length, head width, eye separation, ear separation, ear length, nose length, nose width, leg length, vertical extent, and horizontal extent. Networks were trained for a fixed 250 epochs per pair of stimuli. This was done to reflect the fact that in the Quinn and Eimas studies, infants were shown pictures for a fixed duration of time.

Twelve items from one category were presented sequentially to the network in groups of two (i.e., weights were updated in batches of two) to capture the fact that pairs of pictures were presented to the infants during the familiarization trials. The remaining six items from each category were used to test whether the networks had formed categorical representations.

Like infants, these networks form both Cat and Dog categories. Fig. 1 shows what happens when networks trained on cats are presented with a novel cat and a dog, and when networks trained on dogs are tested with a novel dog and a cat. In this model, higher error is taken to correlate with longer looking time, as it will take a greater number of iterations to reduce the error: the infant will be interested in an item until the system has assimilated it. When the networks are initially trained on cats, the presentation of a novel dog results in a large error score compared with a novel cat, corresponding to the results observed with infants in terms of a longer looking time. Dogs are not included within the category representation of cats. In contrast, when the networks are initially trained on dogs, the presentation of a novel cat results in only a small increase in error compared with a novel dog suggesting that the cats have been included in the dog category.

One advantage of building a model is that it can be taken apart to explore what causes the observed behaviors. Connectionist
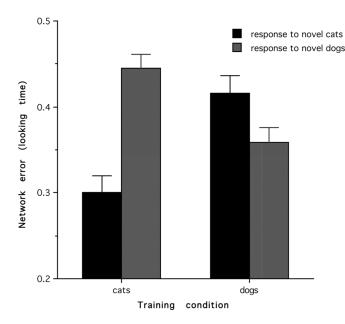
Fig. 1. Network error (predicted looking time) to novel cat and novel dog stimuli for networks trained on (a) cats and (b) dogs (from [44] reprinted with permission).

networks extract the correlations between features present in their learning environment. The variation of the internal representations (developed across the hidden units) reflects the variation of the corresponding categories in the environment. Fig. 2 shows the frequency distributions of the ten input features for both cats and dogs. Each feature has been fit to a normal distribution. In almost all cases, the distribution for each Dog trait (represented by the thick dark lines) subsumes the distribution for the corresponding trait for cats. The narrower distributions for most Cat traits (represented by the thin light lines), on the other hand, do not subsume the range of values for the corresponding Dog traits. In other words, cats are possible dogs but the reverse is not the case: most dogs are not possible cats.

The crucial distributional feature of the data is that Cat features are (in general) subsumed within the distribution of Dog features. It is not just the added variability of dogs along certain features, but the subset relationship that is crucial for explaining the asymmetry. Connectionist networks develop internal representations that reflect the distributions of the input features. Thus, the internal representation for Cat will be subsumed within the internal representation for Dog. It is because the internal representations share this inclusion relationship that an asymmetry in error (looking time) is observed. The behavior arises because of an interaction between the statistics of the environment and the computational properties of the learning algorithm.

Because of its close tie to the original experiments, the same model can then be used to explain and predict infant behaviors in novel contexts. If the infant looking time behaviors are indeed driven by the inclusivity relation among the distribution of features in the sets of pictures shown to the infants during familiarization, then it should be possible to manipulate their looking time by manipulating the distributions of features. This is exactly what was found. French and colleagues [25] developed sets of images that either inverted the inclusion relations
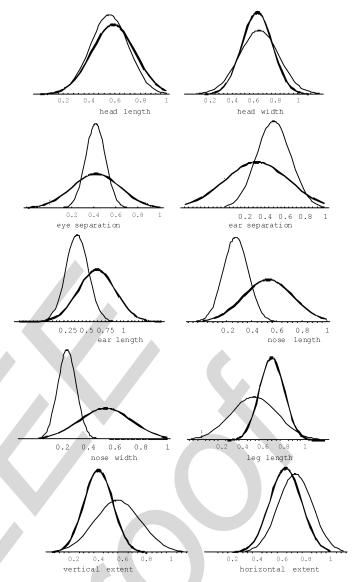

Fig. 2. Normal distributions of cat and dog feature values generated from the means and standard deviations of measurements taken from stimuli used to test infants (from [44] reprinted with permission).

(Dogs were now included within Cats) or removed the inclusion relations (neither set was included within the other). When infants were familiarized with these sets of images in the same way as before, they then showed either a reversed looking time pattern or complete separation, respectively.

In summary, this model illustrates how categorical representations of visually presented stimuli can be acquired within a testing session. An associative system that parses stimuli into distinct features and develops distributed representations will also develop categories with the same exclusivity asymmetries as 3- to 4-month-olds when presented with the same stimuli as these infants. This analysis constitutes a novel explanation of the infant data that emerges from the construction of a computational model that makes *testable* empirical predictions.

B. The Emergence of Correlation-Based Category Learning

One striking aspect of infant categorization is a shift in the way in which infants process objects with increasing age.

Younger and Cohen [116] have argued that while 4-month-olds process objects globally (i.e., as the sum of their individual features), 10-month-olds take correlations between features into account when forming categories. Experiments by Younger and Cohen [115], [116] employed a looking-time procedure in which infants were shown line drawings of artificial cartoon animals. In these experiments, infants were familiarized to a number of stimuli that could be construed as forming one or several categories. Younger found that when a test item violated the correlations between features that were present in the training items, 10-month-olds but not 4-month-olds, noticed this difference. This shift in categorization behavior from 4- to 10-month-olds was claimed to be qualitative, from processing features in isolation to processing relations between features. Features in the real world are correlated and, thus, the ability to process combinations between them provides valuable information about categories.

In contrast, Westermann and Mareschal [109] argue that the change in infant categorization behavior observed between 4 and 10 months of age is caused by a gradual refinement of the infants' representational acuity. This hypothesis, called the *Representational Acuity Hypothesis* can be directly linked to aspects of neural development in the cortex. The main assumption of the Westermann and Mareschal hypothesis is that the receptive field size of units on the representational map decrease during development, and this decrease leads to an increase in representational acuity. This process is analogous to that which occurs in the development of visual acuity. During the first months of life, increases in visual acuity are believed to be partly dependent on the decrease of receptive field size of visual neurons ([83], Wilson, 1988). In the case of representational acuity, the authors argue that the decrease of receptive field sizes of neurons in a higher cortical area has a similar functional consequence: it leads to an increased accuracy in the feature-based representation of objects and causes the developmental shift in categorization behavior from global to local processing of objects.

This hypothesis was tested with simulations using a Gaussian autoencoder network (an autoencoder network whose hidden units have Gaussian activation functions with localized receptive fields). The networks were trained and tested using the exact same regime and animal stimuli as Younger and Cohen [116] used to test the infants. To mimic the infant testing procedure, the model was trained on each familiarization item (animal picture) for a fixed number of learning steps, and after all training items had been presented, it was tested on the three items used to test the infants (correlated features, uncorrelated features, and novel). Once again, network output error was taken as a measure of the looking time towards a particular item.

The result of this simulation is shown in Fig. 3. With large receptive fields, the errors for the correlated and uncorrelated animal stimuli were the same, whereas the error for the novel stimulus was higher, reflecting the profile of 4-month-olds. Again, with decreasing receptive fields, the novel error decreased, whereas the uncorrelated error increased, leading to the profile observed in 10-month-olds. A closer look at the simulation results showed that the model also captured fine details of the infants' behavior such as the decrease in looking time towards the novel stimulus with increasing age.
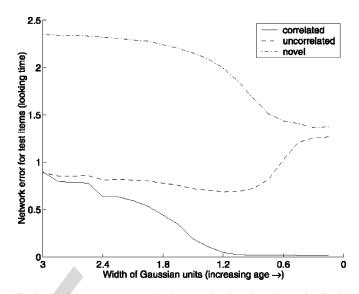

Fig. 3.   Model response to correlated, uncorrelated, and novel test stimuli with increasing age (decreasing receptive field sizes) (from [109] reprinted with permission).

In summary, the Gaussian autoencoder model is able to capture both the main effects observed across development, as well as the finer grained behavioral detail in the infant data. This simulation illustrates how shrinking receptive fields could explain the developmental profile of infants even when not all feature values in the stimuli were correlated. The strength of the Gaussian autoencoder model is that: 1) it relies on general mechanisms known to function in other domains; 2) it links neural and cognitive levels of description; and 3) because it is clearly tied to empirical experiments, it can be used to make specific testable prediction of novel behaviors. While Younger and Cohen [116] postulated that the infant undergoes a qualitative shift from processing features to a mode of processing correlations between features, the model processes correlations between features from the start. The changes in behavior are produced by increases in representational acuity. Thus, the model produces an alternative account to that offered by Younger and Cohen (see [96], for discussion). The notion that changes in representational acuity may explain developmental shifts in behavior has found wider application, for example, in work on semantic development [79].

C. Specific Language Impairment (SLI) and Inflectional Morphology

SLI is a developmental disorder of language in the absence of evidence for brain damage, hearing deficits, severe environmental deprivation, or general learning disabilities [41]. Nonverbal abilities are usually in the normal range, but its absolute specificity to language is a matter of debate (see [106], for a recent review). SLI is a heterogeneous disorder that may impact on different facets of language to different extents, and behavioral genetics indicates that the disorder has a strong genetic component [5]. Children with SLI frequently exhibit particular problems in morphology and in syntax. For example, Rice [75] has argued that for English, problems in acquiring the English past tense may be a useful diagnostic for the disorder.
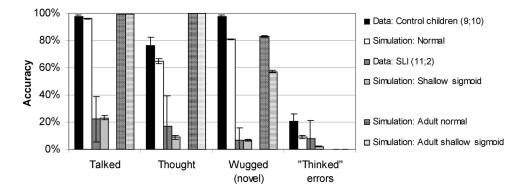
Fig. 4. Accuracy levels in an English past tense elicitation task for typically developing children (mean age 9 years, 10 months; [103]), children with SLI (mean age 11 years, 2 months; [108]), and the model trained under normal conditions or with a shallow sigmoid in the activation function of its processing units (see [97]). The model is trained for 250 epochs for the child data and 500 epochs for adult outcome.

The English past tense is marked by a distinction between regular verbs (e.g., talk-talked) and irregular verbs (e.g., go-went, think-thought, hit-hit), and the regular pattern is frequently extended by children to novel words (e.g., wug-wugged). In SLI, children exhibit low levels of accuracy in producing the English past tense, both for regulars and irregulars, instead producing predominantly unmarked forms (e.g., talk) in contexts where the past tense is appropriate. Fig. 4 includes representative empirical data for a sample of children with SLI age around 11 years (from [108]), compared with a slightly younger control group (from [103]). Regulars are produced at an accuracy level of about 20% but still hold a slight advantage over irregulars as in the control children. There are very much reduced levels of generalization to novel stems (e.g., wugged) in the SLI group, but still sparse overgeneralization errors (e.g., thinked). Since younger typically developing children go through a stage of not inflecting verbs, these low levels of accuracy are sometimes seen as an atypical extension of a normal "optional infinitive" stage of language development (see [75]).

In comparison to the control sample, the children with SLI suffer a relatively larger deficit on regular verbs than on irregular verbs. Ullman and Pierpont [106] have recently proposed that the SLI pattern may be explained in terms of a developmental problem with the part of the brain that learns automatic skills, which they implicate in learning the regular past tense rule. The residual level of past tense performance is then taken to result from compensation from a separate lexical memory system (see also [108]). That is, these children cannot learn the rule and instead just learn a set of individual past tenses by rote. Evidence for this comes from an exaggeration of frequency effects in regular verb formation in SLI. Typically, developing children do not show this, and frequency is taken to be a hallmark of a memory system not a skill system.

A large body of work using connectionist networks to study normal development of the English past tense already exists (see [101] for a review). These models are not uncontested in the field [43], [65], but few alternate accounts have been explored computationally (see [98]). The presence of a normal model allows us to explore parameters that could produce various atypical profiles. The normal model (e.g., [35]) sees this skill as requiring the individual to learn to modulate the sound of the verb in order to fit the intended grammatical context of the sentence, based on knowledge of the phonological form of the verbs stem
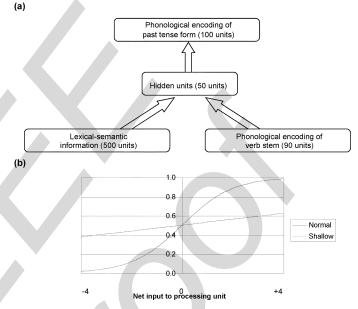

Fig. 5. (a) Past tense model architecture; the model is a backpropagation feed-forward network. (b) Manipulation to the activation function of the hidden and output units, applied to the start state of the atypical model.

and input from lexical semantics about that word's identity. This architecture is shown in Fig. 5(a).

Thomas and Karmiloff–Smith [101] assessed the result of a wide range of manipulations to this normal model in order to explore the developmental parameter space of the system, that is, what range of atypical developmental trajectories and error patterns could be produced under different constraints. One of the parameter manipulations produced a reasonable fit to the empirical data for SLI [97], shown in Fig. 4 (see also [34] and [36], for other related approaches to modeling SLI). The model captured the low level of inflections, a greater deficit for regulars than irregulars, the slight advantage for regulars over irregulars, and the low residual levels of generalization and overgeneralization errors. This trajectory was the result of reducing the slope of the sigmoid activation function in the processing units comprising the hidden and output layers [shown in Fig. 5(b)]. Three interesting conclusions can be drawn from this result.

First, the shallow sigmoid produced a problem in learning because it impaired the formation of sharp category boundaries

to partition the internal representational space of the network. Sharp boundaries are necessary to learn generalizable regularities: every item that falls inside the category should be treated the same. The suboptimal activation function delayed all learning, but retarded the regular verbs more than the irregular verbs despite the fact that both were processed in a common channel. This contrasts with the inference drawn by Ullman and Pierpont [106] that the SLI profile must imply damage to a rule-specific structure in the learning system. Here, the processing property of units in the shared channel happened to be more relevant to one class of mappings than another. In other words, domain-specific behavioral deficits do not have to be caused by initial damage to domain-specific processing structures.

Second, in addition to fitting the overall accuracy levels, the atypical model replicated the empirical pattern of exaggerated frequency effects observed in residual regular verb performance [97]. The absence of frequency effects in regulars in normal models was the result of a ceiling effect in the nonlinear processing units for this mapping class. In the backpropagation learning algorithm, weight change is proportional to the slope on the sigmoid activation function. Because the low derivative of the shallow sigmoid slowed down learning, it encouraged item-based effects such as frequency. Furthermore, in the atypical model, the changed internal parameter altered the balance between the information sources used to drive performance. In the normal model, the phonological input tends to drive regular verb performance, while lexical-semantic input tends to drive irregulars. In the atypical model, lexical-semantics drove both regular and irregular verb performance, because the model was struggling to learn the phonological regularities linking the stem and its inflected form. Residual performance was therefore achieved using atypical representations.

Third, with the model we can run on training to see whether this developmental problem with past tense formation resolves itself. Fig. 4 also demonstrates performance at the end of training. The atypical model has now reached ceiling on the training set – it has caught up, as if its development was a form of "delay."[5] However, importantly, the system retained a deficit in generalization to novel stems (wugged). Moreover, in its end state performance, the atypical balance between the sources of information driving the output also persisted. Therefore, although further training eventually produced behavior on the *training set* that looked normal, the network had actually learned an atypical underlying function, revealed by its different use of input information, and by the inductive bias that diverged from the normal case. Relatively little work has explored the extended outcome of SLI into adulthood, but some parallels may be drawn between existing data and the model's prediction. Although by adulthood there is evidence of recovery in some areas of SLI such as vocabulary, persistent problems can be found in generalizing speech sounds to novel stimuli in short-term memory tasks such as being asked to repeated nonsense words [6], [104].

[5] It can be shown formally that reducing the slope on the sigmoid can be compensated for by increasing the size of the weights feeding into a processing unit; additional training serves to increase weight size and so can ameliorate the effects of the shallow slope.

The model of SLI has some limitations. As a result of its simplified architecture, it simulates accuracy levels rather than error patterns in past tense, and its normal level of generalization is lower than that shown by children. However, the simulation: 1) captured atypical accuracy patterns as the outcome of a developmental process; 2) also simulated atypical frequency effects in past tense production observed in children with SLI; and 3) generated testable predictions about resolution of the disorder. Finally, it suggested that mechanisms explaining the behavioral impairments in the disorder might not be specific to the domain of past tense (or even language), but rather a general property of associative systems that is atypical in the inflectional system. This account contrasts with domain-specific explanations of SLI, for example, that these children have an "extended optional infinitive" stage in their language development [75].

D. The Development of Gaze Following Behavior in Infants With Autism and Williams Syndrome

The previous example discovered a mechanistic explanation for a case of disordered development via an exploration of the developmental parameter space of a normal model. Parameter manipulations can, of course, be principled and based on known features of the disorder. The implementation then serves to test whether one feature of the target disorder may serve as a causal explanation for other observed features of the disorder via a developmental process. The following model sought to assess this possibility for two disorders, autism and Williams syndrome, with respect to a single target behavior of eye gaze. Note, of course, that the model was not intended as a general account of either disorder: both autism and Williams syndrome have multifaceted atypical cognitive profiles).

Triesch *et al.* [105] recently proposed a computational model of the emergence of gaze following skills in infant-caregiver interactions. Shared attention plays a crucial role in the communication between infants and their caregivers and by 9–12 months, most infants can follow an adult's gaze and pointing gestures to locate objects in the world. This ability might be thought to require a level of social understanding: the infant realizes that the adult intends to refer to some event in the world. In contrast, Triesch *et al.* constructed their model to test the idea that the emergence of gaze following may be explained more simply as the infant's gradual discovery that monitoring the caregiver's direction of gaze is predictive of where rewarding objects are located. In addition to capturing normal development, Triesch *et al.* then sought to capture the eye gaze profiles demonstrated by two developmental disorders. In autism, deficits in shared attention are a consistent early predictor of subsequent social and language deficits [62]. In Williams syndrome (WS), a rare genetic disorder [16], deficits in shared attention have been observed in toddlers despite a hypersocial personality profile [40].

In contrast to the backpropagation network models used in the previous examples, Triesch *et al.* based their model of gaze following on a biologically plausible reward-driven mechanism called Temporal Difference learning (a type of reinforcement learning). The infant is construed as an agent situated in an environment. The agent generates actions based on what it perceives from the environment, and then potentially receives a reward for its action along with updated information of the new

state of the environment. In the current model, the environment depicted a range of locations that might contain the caregiver, an interesting object, or nothing. If the infant were fixating the caregiver, information would also be available on the direction of the caregiver's gaze (i.e., at the infant or to some location in the environment). Rewards were available to the infant for fixating an object or the caregiver, but these rewards habituated (reduced) over time as the infant became bored.

The aim of temporal difference learning is to acquire estimates of the potential reward of each of the system's actions in response to each possible environmental state it might encounter. It achieves this initially by exploring the environment to see what rewards are available. Note that some rewards are gained by sequences of actions, and so reward estimate information must percolate back in time from when the reward occurred to the sequence of actions that preceded it. For example, if fixating an object generates a reward and that action happens to follow a previous action of choosing the location indicated by the caregiver's gaze, then this prior action should also have its reward estimate increased. Similarly, this sequence must have begun by fixating the caregiver, so again, this first action should have its reward estimate increased. The system faces two dilemmas: 1) Should it *explore* the environment to gain more information about where the greatest rewards lie or should it *exploit* its current knowledge to maximize rewards? (2) Should the system accept small rewards generated by an immediate action, or should it defer in favor of an alternative action that may produce a lesser reward now but potentially a larger reward in the longer term? Both dilemmas correspond to algorithmic variables, some of the free parameters in the model.

The Triesch *et al.* [105] model of gaze following had two components, one to predict *when* it would be advantageous for the infant to shift gaze, and a second to predict *where* the optimal location in the environment would be to shift gaze to (including the caregiver's face and locations of potential objects in the environment). Through rewards gained during exploration of a simulated environment, seeded with interesting objects in fixed random locations and a caregiver who would sometimes look at these objects (and sometimes at the infant), the model successfully acquired gaze following. Not unexpectedly, the ability to shift gaze to look at the caregiver preceded the ability to use the caregiver's gaze direction as a cue to fixate the relevant location.

With the normal model in hand, Triesch *et al.* then used the following facts from the two developmental disorders. In autism, children frequently show gaze avoidance, not looking at the faces of other people and particularly avoiding their eyes (e.g., [15]). It is possible children with autism find eyes to be aversive stimuli rather than simply neutral. In WS, the converse is the case. Infants and children with WS find faces fascinating, often fixating on them for long periods in preference to other objects. This is part of a personality profile that has been described as hypersocial (e.g., [3] and [37]). Triesch *et al.* used these constraints to alter the reward contingencies for the two disorders: for autism, fixating objects was interesting but fixating the caregiver's face was given a negative reward; for WS, fixating faces was given a much higher reward. In both cases, the model generated an atypical developmental trajectory, with the emergence of gaze following absent or substantially delayed.

Given the implemented model, it is also possible to generate predictions on how gaze following might be affected in other disorders. Attention Deficit Hyperactivity Disorder (ADHD) is a developmental disorder that is not typically diagnosed until childhood. However, its basis appears to be in part genetic (e.g., [2]). Using a Temporal Difference learning model, Williams and Dayan ([112]; see also [113]) have shown that the behavioral features of ADHD may be simulated by altering the parameters of the system so that larger long-term rewards are discounted in favor of smaller short-term rewards: the system is effectively impulsive and unable to pursue long-term goals. Richardson and Thomas [77] demonstrated that if the parameter manipulations of the Williams and Dayan ADHD model are applied to Triesch *et al.*'s model, then impairments in the development of early gaze following are also observed.[6] If the genetic influence on ADHD means that precursors to the childhood behavioral symptoms can be observed in infancy, this modeling result predicts that atypical gaze following may be just such a precursor. Importantly, this precursor pattern would not match autism (in which the infant avoids the caregiver's face) or WS (in which there is longer fixation on the caregiver's face) but pertain specifically to reduced use of the caregiver's face to predict object locations, compared to noncontingent gaze shifts to the caregiver's face or to objects.

Our second disorder example again builds on the initial implementation of a normal model of development. In this case: 1) two empirically motivated, *a priori* manipulations of start state parameters (differential rewards for certain actions) were used to test the validity of causal theories of how behavioral features in two disorders may be linked across development (gaze avoidance and atypical shared attention, exaggerated interest in faces and atypical shared attention, respectively) and 2) novel predictions were generated regarding potential prediagnosis precursors in a further disorder.

Taken together, the use of the developmental computational models to investigate developmental disorders underscores a theoretical point made by Karmiloff–Smith [38]. Disorders that appear very different in their adult states may, in fact, be traced back to infant systems that share much in common, but differ in certain low-level neurocomputational properties (see [45], for further discussion). It is development itself – together with the characteristics of the system that is undergoing development – that produces divergent behavioral profiles.

IV. CONCLUSION

Computational modeling can make a fundamental contribution to the understanding of perceptual and cognitive development. The development of relevant computational models is an essential step in moving from a descriptive science that just tracks what changes occur over time, to a causal science that proposes to explain the emerging behaviors that are observed in terms of well specified dynamic processes. Because they are well specified, the candidate causal processes and mechanisms can also be related to neural processes (if necessary), thus enabling a bridge between multiple levels of descriptions to be

---

[6]Specifically, in separate conditions, Richardson and Thomas altered the learning rate, temperature, and discounting rate parameters of the temporal difference algorithm, in line with Williams and Dayan's [112] proposals of possible ways to capture the ADHD profile.

built. Transitions from descriptive to explanatory sciences have already occurred in most of the physical sciences. The time is more than ripe for such a transition in developmental psychology.

We end our article with a cautionary note. It is essential to distinguish between the use of computational models in a psychological/explanatory mode, and their use in an engineering mode. As we argued in the introduction, these approaches can be complementary and can cross-fertilize. The study of biological systems can generate new possible engineering solutions. Engineering approaches to machine learning and robotics – unfettered by psychological constraints and required to build systems that respond in real time given the noisy data typically furnished by the real world – can reveal new ways in which algorithmic systems can learn. However, we sincerely hope modelers using both approaches would focus on explanation, as well as implementational issues. The key to success is not just getting the models to work but in understanding why they work. If the community cannot understand how the models work, then (however admirable the achievement) the models have little explanatory value. In this context, it is worth considering a line by the comic science fiction writer Douglas Adams brought to our attention by Andy Clark in his book *Microcognition*:

> The Hitch Hiker's guide to the Galaxy, in a moment of reasoned lucidity which is almost unique amongst its current tally of 5,975,509 pages, says of the Sirius Cybernetics Corporation products that "it is very easy to be blinded to the essential uselessness of them by the sense of achievement you get from getting them to work at all. In other words – and this is the rock solid principle on which the whole of the corporation's Galaxy-wide success is grounded – their fundamental design flaws are completely hidden by their superficial design flaws" [13, p. 7].

A model of development that ignores the scientists who will use it and build upon it runs the risk of sharing many features with the *Sirius Cybernetics Corporation* products. Fortunately, this does not need to be the case and, indeed, has not been the case for many highly successful published models of development.

REFERENCES

[1] K. Ahmed, M. Casey, and T. Bale, "Connectionist simulation of quantification skills," *Connection Sci.*, vol. 14, pp. 165–201, 2002.

[2] T. Banaschewski, C. Hollis, J. Oosterlaan, H. Roeyers, K. Rubia, E. Willcutt, and E. Taylor, "Towards an understanding of unique and shared pathways in the psychopathophysiology of ADHD," *Developmental Sci.*, vol. 8, pp. 132–140, 2005.

[3] U. Bellugi, L. Lichtenberger, W. Jones, Z. Lai, and M. St. George, "The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses," *J. Cogn. Neurosci.*, vol. 12, pp. 1–29, 2000.

[4] J. B. Birks, *Rutherford at Manchester*.   Manchester, U.K.: Heywood and Co., 1962.

[5] D. V. M. Bishop, T. North, and C. Donlan, "Genetic basis of specific language impairment," *Developmental Medicine & Child Neurology*, vol. 37, pp. 56–71, 1995.

[6] ——, "Nonword repetition as a behavioral marker for inherited language impairment: Evidence from a twin study," *J. Child Psychol. Psych. Allied Disciplines*, vol. 37, pp. 391–403, 1996.

[7] M. A. Boden, *Computer Models of Mind*.   Cambridge, U.K.: Cambridge Univ. Press, 1988.

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

[9] J. Changeux and S. Dehaene, "Neural models of cognitive functions," *Cognition*, vol. 33, pp. 63–109, 1989.

[10] M. H. Christiansen, J. Allen, and M. S. Seidenberg, "Learning to segment speech using multiple cues: A connectionist model," *Language and Cognitive Processes*, vol. 13, pp. 221–268, 1998.

[11] M. H. Christiansen and N. Chater, Eds., Connectionist Psycholinguistics   Westport, CT, Ablex, 2001a.

[12] M. H. Christiansen and N. Chater, "Connectionist psycholinguistics: Capturing the empirical data," *Trends in Cognitive Sciences*, vol. 5, pp. 82–88, 2001b.

[13] A. Clark, *Microcognition*.   Cambridge, MA: MIT Press.

[14] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control, Signals, Syst.*, vol. 2, pp. 303–314, 1989.

[15] G. Dawson, A. N. Meltzoff, J. Osterling, J. Rinaldi, and E. Brown, "Children with autism fail to orient to naturally occurring social stimuli," *J. Autism and Developmental Disorders*, vol. 28, pp. 479–485, 1998.

[16] D. Donnai and A. Karmiloff-Smith, "Williams syndrome: From genotype through to the cognitive phenotype," *Amer. J. Medical Genetics*, vol. 97, pp. 164–71, 2000.

[17] G. L. Drescher, *Made-Up Minds. A Constructivist Approach to Artificial Intelligence*.   Cambridge, MA: MIT Press, 1991.

[18] J. L. Elman, "Learning and development in neural networks: The importance of starting small," *Cognition*, vol. 48, pp. 71–99, 1993.

[19] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett, *Rethinking Innateness: A Connectionist Perspective on Development*.   Cambridge, MA: MIT Press, 1996.

[20] R. L. Fantz, "Visual experience in infants: Decreased attention to familiar patterns relative to novel ones," *Science*, vol. 164, pp. 668–670, 1964.

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

[22] ——, "Wavelets, vision and the statistics of natural scenes," *Philos. Trans. Roy. Soc. London A*, vol. 357, pp. 2527–2542, 1999b.

[23] J. H. Flavell, *The Developmental Psychology of Jean Piaget*.   New York: Van Nostrand, 1963.

[24] N. Forrester and K. Plunkett, A. Ramand and K. Eiselt, Eds., "Learning the Arabic plural: The case for minority default mappings in connectionist networks," in *Proc. Sixteenth Annu. Conf. Cogn. Sci. Soc.*, 1994, pp. 319–323.

[25] 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," *J. Experimental Psychology: General*, vol. 133, pp. 382–397, 2004.

[26] M. Gasser and L. B. Smith, "Learning nouns and adjectives: A connectionist account," *Language and Cognitive Processes*, vol. 13, pp. 269–306, 1998.

[27] S. Grossberg and D. Seidman, "Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates," *Psychol. Rev.*, vol. 113, pp. 483–525, 2006.

[28] U. Hahn and R. C. Nakisa, "German inflection: Single or dual route?," *Cogn. Psych.*, vol. 41, pp. 313–360, 2000.

[29] M. Harm and M. S. Seidenberg, "Phonology, reading acquisition, and dyslexia: Insights from connectionist models," *Psychol. Rev.*, vol. 106, pp. 491–528, 1999.

[30] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2 ed.   London, U.K.: Prentice-Hall International, 1999.

[31] J. Hertz, A. Krogh, and R. G. Palmers, *Introduction to the Theory of Neural Computation*.   Readintg, MA: Addison-Wesley, 1991.

[32] W. F. Hill, *Learning: A Survey of Psychological Interpretations*.   London, U.K.: Chandler, 1967.

[33] G. E. Hinton, "Connectionist learning procedures," *Artif. Intell.*, vol. 40, pp. 185–234, 1989.

[34] J. H. Hoeffner and J. L. McClelland, E. V. Clark, Ed., "Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation," in *Proc. 25th Child Language Research Forum*, Stanford, CA, 1993, pp. 38–49.

[35] M. F. Joanisse and M. S. Seidenberg, "Impairments in verb morphology following brain injury: A connectionist model," in *Proc. Nat. Acad. Sci.*, USA, 1999, vol. 96, pp. 7592–7597.

[36] ——, "Phonology and syntax in specific language impairments: Evidence from a connectionist model," *Brain and Language*, vol. 86, pp. 40–56, 2003.

[37] W. Jones, U. Bellugi, Z. Lai, M. Chiles, J. Reilly, A. Lincoln, and R. Adolphs, "Hypersociability in Williams syndrome," *J. Cogn. Neurosci.*, vol. 12, pp. 30–46, 2000, Supplement.

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

[39] D. Klahr and J. G. Wallace, *Cognitive Development: An Information Processing View*. Mahwah, NJ: Erlbaum, 1976.

[40] E. Laing, G. Butterworth, D. Ansari, M. Gsödl, E. Laing, Z. Barnham, L. Lakusta, L. K. Tyler, S. Grice, S. Paterson, and A. Karmiloff-Smith, "Atypical linguistic and socio-communicative development in toddlers with Williams syndrome," *Developmental Sci.*, vol. 5, no. 2, pp. 233–246, 2002.

[41] L. B. Leonard, *Children with Specific Language Impairment*. Cambridge, MA: MIT Press, 1998.

[42] S. Lewandowsky, "The rewards and hazards of computer simulations," *Psych. Sci.*, vol. 4, pp. 236–243, 1993.

[43] G. F. Marcus, *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. Cambridge, MA: MIT Press, 2001.

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

[45] D. Mareschal, M. H. Johnson, S. Sirois, M. Spratling, M. Thomas, and G. Westermann, *Neuroconstuctivism*. Oxford, U.K.: OUP, 2007, vol. 1, How the brain constructs cognition.

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

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

[48] D. Mareschal and T. R. Shultz, "Generative connectionist networks and constructivist cognitive development," *Cognitive Development*, vol. 11, pp. 571–604, 1996.

[49] ——, "Development of children's seriation: A connectionist approach," *Connection Sci.*, vol. 11, pp. 153–188, 1999.

[50] D. Mareschal, S. Sirois, Westermann, and M. H. Johnson, *Neuroconstructivism*. Oxford, U.K.: Oxford Univ. Press, 2007, vol. 2, Perspectives and Prospects.

[51] J. L. McClelland and E. Jenkins, "Nature, nurture, and connections: Implications of connectionist models for cognitive development," in *Architectures for Intelligence*, K. van Lehn, Ed. Hillsdale, NJ: LEA, 1991, pp. 41–73.

[52] D. McDermott, "Artificial intelligence meets natural stupidity," in *Mind Design: Philosophy, Psychology, Artificial Intelligence*, J. Haugel, Ed. Cambridge, MA: MIT Press, 1981.

[53] J. L. McDonald, "Language acquisition: The acquisition of linguistic structure in normal and special populations," *Annu. Rev.Psych.*, vol. 48, pp. 215–241, 1997.

[54] P. McLeod, K. Plunkett, and E. T. Rolls, *Introduction to Connectionist Modeling of Cognitive Processes*. Oxford, U.K.: Oxford Univ. Press, 1998.

[55] T. M. Mitchell, *Machine Learning*. McGraw-Hill, New York: , 1997.

[56] Y. Munakata, "Infant perseveration and implications for object permanence theories: A PDP model of the A-not-B task," *Developmental Sci.*, vol. 1, pp. 161–211, 1998.

[57] Y. Munakata and J. L. McClelland, "Connectionist models of development," *Developmental Sci.*, vol. 6, pp. 413–429, 2003.

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

[59] R. C. Nakisa and K. Plunkett, "Innately guided learning by a neural network: The case of featural representation of speech language," *Language and Cogn. Processes*, vol. 13, pp. 105–128, 1998.

[60] A. Newell, *Unified Theories of Cognition*. Cambridge, MA: Harvard Univ. Press, 1990.

[61] Y. Oshima-Takane, Y. Takane, and T. Shultz, "The learning of first and second person pronouns in English: Network models and analysis," *J. Child Language*, vol. 25, pp. 545–575, 1999.

[62] J. Osterling and G. Dawson, "Early recognition of children with autism: A study of first birthday home video tapes," *J. Autism and Developmental Disorders*, vol. 24, pp. 247–257, 1994.

[63] J. Piaget, *Problemes de Psychologie Genetique*. Paris, France: Editions Denoel, 1972.

[64] ——, *Le Psychologie de L'intelligence*. Paris, France: Armad Collin, 1977.

[65] S. Pinker, *Words and Rules*. London, U.K.: Weidenfeld Nicolson, 1999.

[66] 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," *Psychol. Rev.*, vol. 103, pp. 56–115, 1996.

[67] K. Plunkett and P. Juola, "A connectionist model of English past tense and plural morphology," *Cogn. Sci.*, vol. 23, pp. 463–490, 1999.

[68] K. Plunkett and V. Marchman, "U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition," *Cognition*, vol. 38, pp. 1–60, 1991.

[69] ——, "Learning from a connectionist model of the acquisition of the English past tense," *Cognition*, vol. 61, pp. 299–308, 1996.

[70] K. Plunkett and R. C. Nakisa, "A connectionist model of the Arabic plural system," *Language and Cognitive Processes*, vol. 12, pp. 807–836, 1997.

[71] K. Plunkett and C. Sinha, "Connectionism and developmental theory," *Br. J. Developmental Psych.*, vol. 10, pp. 209–254, 1992.

[72] K. Plunkett, C. Sinha, M. F. Moller, and O. Strandsby, "Vocabulary growth in children and a connectionist net," *Connection Sci.*, vol. 4, no. 3-4, pp. 293–312, 1992.

[73] P. C. Quinn, P. D. Eimas, and S. L. Rosenkrantz, "Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants," *Perception*, vol. 22, pp. 463–475, 1993.

[74] M. E. J. Raijmakers and P. C. M. Molenaar, "Modeling developmental transitions in adaptive resonance theory," *Developmental Sci.*, vol. 7, pp. 149–157, 2004.

[75] M. Rice, "Grammatical symptoms of specific language impairment," in *Speech and Language Impairments in Children*, D. Bishop and L. Leonard, Eds. Hove, Sussex: Psychology Press, 2000, pp. 17–34.

[76] F. M. Richardson, F. D. Baughman, N. A. Forrester, and M. S. C. Thomas, "Computational modeling of variability in the balance scale task," in *Proc. 7th Inte. Conf. Cogn. Modeling*, Trieste, Italy, 2006, pp. 256–261, Trieste, Italy: Edizioni Goliardiche.

[77] F. Richardson and M. S. C. Thomas, "The benefits of computational modeling for the study of developmental disorders: Extending the Triesch et. al. model to ADHD," *Developmental Sci.*, vol. 9, pp. 151–155, 2006.

[78] M. Robert and A. Galatzer-Levy, "Chaotic possibilities: Toward a new model of development," *Int. J. Psychoanalysis*, vol. 85, pp. 419–441, 2004.

[79] T. T. Rogers and J. L. McClelland, *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press, 2004.

[80] D. Ruelle, *Chaotic Evolution and Strange Attractors*. Cambridge, U.K.: Cambridge Univ. Press, 1989.

[81] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986a, vol. 1.

[82] ——, "On learning the past tense of English verbs," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* J. L. McClelland and D. E. Rumelhart, Eds. Cambridge, MA: MIT Press, 1986b, vol. 2, Psychological and Biological Models, pp. 216–271.

[83] A. Rusoff and M. Dubin, "Development of receptive field properties of retinal ganglion cells in kittens," *J. Neurophysiology*, vol. 40, pp. 1188–1198, 1977.

[84] G. Schafer and D. Mareschal, "Modeling infant speech sound discrimination using simple associative networks," *Infancy*, vol. 2, pp. 7–28, 2001.

[85] M. Schlesinger, "Evolving agents as a metaphor for the developing child," *Developmental Sci.*, vol. 7, pp. 154–168, 2004.

[86] W. C. Schmidt and X. S. Ling, "A decision-tree model of balance scale development," *Mach. Learn.*, vol. 24, pp. 203–230, 1996.

[87] M. S. Seidenberg and J. L. McClelland, "A distributed, developmental model of word recognition and naming," *Psychol. Rev.*, vol. 96, pp. 452–477, 1989.

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

[89] ——, *Computational Developmental Psychology*. Cambridge, MA: MIT Press, 2003.

[90] T. R. Shultz, D. Buckingham, and Y. Oshima-Takane, "A connectionist model of the learning of personal pronouns in English," in *Computational Learning Theory and Natural Learning Systems: Intersection Between Theory and Experiment*, S. J. Hanson, T. Petsche, M. Kearns, and R. L. Rivest, Eds. Cambridge, MA: MIT Press, 1994, vol. 2, pp. 347–362.

[91] T. R. Shultz, D. Mareschal, and W. C. Schmidt, "Modeling cognitive development on balance scale phenomena," *Mach. Learn.*, vol. 16, pp. 57–86, 1994.

[92] H. A. Simon and A. Newell, "Heuristic problem solving: The next advance in operations research," *Oper. Res.*, vol. 6, pp. 1–10, 1958.

[93] S. Sirois, "Autoassociator networks and insights into infancy," *Developmental Sci.*, vol. 7, pp. 133–140, 2004.

[94] S. Sirois and T. R. Shultz, "Neural network modeling of developmental effects in discrimination shifts," *J. Experimental Child Psychology*, vol. 71, pp. 235–274, 1998.

[95] M. S. C. Thomas, "Multiple causality in developmental disorders: Methodological implications from computational modeling," *Developmental Sci.*, vol. 6, pp. 537–556, 2003.

[96] ——, "How do simple connectionist networks achieve a shift from "featural" to "correlational" processing in categorization?," *Infancy*, vol. 5, no. 2, pp. 199–207, 2004.

[97] ——, "Characterising compensation," *Cortex*, vol. 41, no. 3, pp. 434–442, 2005.

[98] M. S. C. Thomas, N. A. Forrester, and F. M. Richardson, "What is modularity good for?," in *Proc. 28th Annual Conference of the Cognitive Science Society*, Vancouver, BC, Canada, Jul. 26–29, 2006.

[99] M. S. C. Thomas, J. Grant, Z. Barham, M. Gsödl, E. Laing, L. Lakusta, L. K. Tyler, S. Grice, S. Paterson, and A. Karmiloff-Smith, "Past tense formation in Williams syndrome," *Language and Cognitive Processes*, vol. 16, pp. 143–176, 2001.

[100] M. S. C. Thomas and A. Karmiloff-Smith, "Are developmental disorders like cases of adult brain damage? implications from connectionist modeling," *Behav. Brain Sci.*, vol. 25, no. 6, pp. 727–788, 2002.

[101] ——, "Modeling language acquisition in atypical phenotypes," *Psychol. Rev.*, vol. 110, no. 4, pp. 647–82, 2003a.

[102] ——, "Connectionist models of development, developmental disorders and individual differences," in *Models of Intelligence: International Perspectives*R. J. Sternberg, J. Lautrey, and T. Lubart, Eds.   : American Psychological Association, 2003b, pp. 133–150.

[103] M. S. C. Thomas, D. Mareschal, and A. Hinds, "A connectionist account of the emergence of the literal-metaphorical-anomalous distinction in young children," in *Proc. 23rd Annu. Meeting Cogn. Sci. Soc.*, 2001.

[104] M. S. C. Thomas, F. Richardson, C. Donlan, and C. Price, "Compensation in the development of atypical language systems: FMRI findings from an adult SLI case study," in *Proc. 1st Annu. Conf. Comput. Cogn. Neurosci.*, Washington, Nov. 2005, pp. 3–4.

[105] J. Triesch, C. Teuscher, G. Deak, and E. Carlson, "Gaze following: Why (not) learn it?," *Developmental Sci.*, vol. 9, pp. 125–147, 2006.

[106] M. T. Ullman and E. I. Pierpont, "Specific language impairment is not specific to language: The procedural deficit hypothesis," *Cortex*, vol. 41, no. 4, pp. 399–433, 2005.

[107] M. T. Ullman and M. Gopnik, "Inflectional morphology in a family with inherited specific language impairment," *Applied Psycholinguistics*, vol. 20, pp. 51–117, 1999.

[108] H. K. J. Van der Lely and M. T. Ullman, "Past tense morphology in specifically language impaired and normally developing children," *Language and Cognitive Processes*, vol. 16, no. 2/3, pp. 178–217, 2001.

[109] G. Westermann and D. Mareschal, "From parts to wholes: Mechanisms of development in infant visual object processing," *Infancy*, vol. 5, pp. 131–151, 2004.

[110] G. Westermann and E. Miranda, "A new model of sensorimotor coupling in the development of speech," *Brain and Language*, vol. 89, pp. 393–400, 2004.

[111] G. Westermann, S. Sirois, T. R. Shultz, and D. Mareschal, "Modeling developmental cognitive neuroscience," *Trends in Cognitive Sciences*, 2006, to be published.

[112] J. O. H. Williams and P. Dayan, "Dopamine, learning, and impulsivity: A biological account of ADHD," *J. Child and Adolescent Psychopharmacology*, vol. 15, no. 2, pp. 160–179, 2005.

[113] J. O. H. Williams and E. Taylor, "Dopamine appetite and cognitive impairment in attention deficit/hyperactivity disorder," *Neural Plasticity*, vol. 11, no. 1, pp. 115–132, 2004.

[114] R. Young, *Seriation by Children: An Artificial Intelligence Analysis of a Piagetian Task*.   Cambridge, MA: Birkhauser, 1976.

[115] B. A. Younger, "The segregation of items into categories by ten-month-old infants," *Child Development*, vol. 56, pp. 1574–1583, 1985.

[116] B. Younger and L. B. Cohen, "Developmental changes in infants' perception of correlation among attributes," *Child Development*, vol. 57, pp. 803–815, 1986.

Denis Mareschal received the B.Sc. degree in physics and theoretical physics from Cambridge University, Cambridge, U.K., the M.A. degree in psychology from McGill University, Montreal, QC, Canada, and the D.Phil. degree from Oxford University, Oxford, U.K.

He is currently a Professor at Birkbeck College, University of London, London, U.K. His research centers on developing mechanistic models of perceptual and cognitive development in infancy and childhood.

Dr. Mareschal received the Marr Prize from the Cognitive Science Society (USA), the Young Investigator Award from the International Society on Infant Studies (USA), and the Margaret Donaldson Prize from the British Psychological Society.

Michael Thomas received the B.Sc. degree in psychology from the University of Exeter, Exeter, U.K., the M.Sc. degree in cognitive science from the University of Birmingham, Birmingham, U.K., and the D.Phil. degree in psychology (on behavioral and computational studies of bilingualism) from the University of Oxford, Oxford, U.K.

He is currently a Reader in Cognitive Neuropsychology at Birkbeck College, University of London, London, U.K. His research focuses on language and cognitive development, and specifically neurocomputational explanations of the variability seen in typical children and in children with developmental disorders (see www.psyc.bbk.ac.uk/research/DNL/).