A Data-Acquisition Model for Learning and Cognitive Development and Its Implications for Autism

Arnon Lotem*
Zoology Dept.
Tel Aviv University
lotem@post.tau.ac.il
http://www.tau.ac.il/∼lotem

Joseph Y. Halpern†
Computer Science Dept.
Cornell University
halpern@cs.cornell.edu
http://www.cs.cornell.edu/home/halpern

Abstract

A data-driven model of learning is proposed, where a network of nodes and links is constructed that represents what has been heard and observed. Autism is viewed as the consequence of a disorder in the data-acquisition component of the model—essentially, it is the result of getting an "inappropriate" distribution of data. The inappropriate data distribution leads to problems in data segmentation, which, in turn leads to a poor network representation. It is shown how the model, given inappropriate data distributions, can reproduce the main cognitive deficits associated with autism, including weak central coherence, impaired theory of mind, and executive dysfunction. In addition, it is shown how the model itself can explain the inappropriate data distribution as the result of an inappropriate initial network. Finally, we discuss the relationships between our model and existing neurological models of autism, and the possible implications of our model for treatment.

Keywords: autism, data-driven model of learning, network, theory of mind, weak central coherence

*Much of this work done while the author was on sabbatical at Cornell University. Cornell's support is gratefully acknowledged. Additional support was provided under grant 353/03 from the Israel Science Foundation.

†Supported in part by NSF under Grant IIS-0090145, a Guggenheim Fellowship, a Fulbright Fellowship, and a grant from the NWO. Sabbatical support from CWI and the Hebrew University of Jerusalem is also gratefully acknowledged.

1 Introduction

Autism is a developmental disorder whose causes are unknown; it is currently defined in terms of its symptoms. These symptoms are grouped into three major categories: deficits in reciprocal social interaction, deficits in verbal and nonverbal communication, and repetitive behaviors and interests [Eigsti and Shapiro 2003; Volkmar and Pauls 2003]. Up until recently, there were three leading theories that attempted to explain the core cognitive deficits in autism, each focusing on a different cluster of symptoms:[1]

- *Theory of mind*, which refers to the ability to infer what others are thinking (believing, desiring) in order to predict their behavior. This is an ability impaired in those suffering from autism [Baron-Cohen 2000]. This cluster of deficits includes difficulty with pretend play, problems in understanding false beliefs, and the inability to tell lies.

- *Weak central coherence*, which refers to the tendency of individual with autism to focus on local details, rather than seeing the big picture [Frith 1989]. For example, autistic individuals seem to have more difficulty than controls in recalling sentences or the gist of a story, while they are as good as controls at recalling unconnected word strings [Hermelin and O'Connor 1967].

- *Executive dysfunction*, which refers to the inability of autistic individuals to be flexible, while still maintaining appropriate plans and inhibitions [Hill 2004; Pennington and Ozonoff 1996; Russell 1997]. This is often manifested in the form of behavior that perseveres inappropriately despite changing goals [Ozonoff 1997].

While identifying these core deficits certainly helps in the study of autism, it does not provide a causal explanation of the disorder, nor does it provide a mechanism explaining how autism emerges. In this paper, we propose a new conceptual model for autism in which the core deficits, and other related symptoms, emerge as a result of a much more basic disorder (which, of course, might itself have a genetic cause or multiple genetic causes). We start by describing an (admittedly naive) model for learning and cognitive development. We then show that, in this model, data-acquisition problems (that may emerge from abnormal sensory input and problems with social orientation) lead to a spectrum of problems similar to those observed in autism. Our model thus explains why primary deficits related to sensory input and social orientation may result in autism and its broader expressions, and why intensive early intervention may sometimes help to remediate autism.

When we refer to data-acquisition problems here, we do not simply mean the somewhat obvious relationship between a lack of data and poor learning. Rather, we argue that if fully-functional processing mechanisms in the brain receive an atypical distribution of data, the result can be the type of abnormal behavior characteristic of those suffering from autism. A simple analogy may help clarify our intuitions. Imagine a robot that is programmed to bake cakes. It expects to find the ingredients for the cake in three containers. Thus, its program says things like "Take 2 cups from container 1, mix it with 1 cup from container 2, wait 10 minutes, add half a cup from container 3, and put in the oven at 400° for one hour." Clearly, if the containers contain the wrong ingredients or insufficient ingredients, then the final product will turn out qualitatively quite different from a "normal" cake, although (indeed, because!) the robot

[1] Some more recent theories, such as the *extreme male-brain theory* [Baron-Cohen 2002], *concept-formation failure* [Snyder et al. 2004], and *underconnectivity* or *temporal binding theories* [Belmonte et al. 2004; Brock et al. 2002; Courchesne and Pierce 2005] are considered later in the paper.

correctly follows the prescribed algorithm. The main problem here is not only that a smaller quantity of ingredients results in a poor cake, but that a mismatch between prescribed quantities and processing time inevitably yields a "different" cake. Many different outputs (cakes) are possible, depending on the ingredients in the containers and their quantity.

In our model, the ingredients in the containers are the data, which include scenes and (sequences of) words, and the robot is the brain's processing algorithm, which has evolved to handle a natural flow of data (at a rate and with a distribution like that faced by an average child). As in our analogy, when the wrong quantities of data are processed by the prescribed algorithms, there is a range of possible abnormal outcomes, depending on the quantity and distribution of data received. The label "autism" actually covers a rather wide spectrum of behaviors. Different types of autism may result from different abnormal distributions of data and rates of data flow. The model suggests that, if we are looking for genetic and biological causes for autism, then we should look at mechanisms that affect data acquisition, or, more precisely, to mechanisms that direct which data is acquired more often and which data should be ignored.

Although our approach has points of connection with a number of other approaches that have appeared in the literature (we discuss the connections in later sections), there seems to be very little work in the literature trying to find a causal model of the type we are looking for, based on problems with underlying mechanisms for data acquisition (although the idea that autism is caused by problems with underlying mechanisms, modeled as computer programs, does appear in very preliminary form in an early paper by Kahn and Arbib [1973]).

The rest of the paper is organized as follows. In Section 2, we discuss our basic framework. In particular, we describe how we assume that a child represents data, using a network consisting of nodes labeled by data and links between them. We describe the key features of the network, how an inappropriate distribution of data could cause differences between the networks of normal children and autistic children, and how these differences could lead to autism. Finally, we discuss evidence in support of the basic model and sketch an explanation within the model itself of how an inappropriate data distribution might arise. In Section 3 we show in more detail how the "autistic" networks of Section 2 can produce many of the classic symptoms associated with impaired theory of mind, weak central coherence, and problems with executive function. In Section 4 we present a computer simulation that illustrates some of the principles of our model and provides additional insight into the model. We conclude in Section 5 with some discussion of the model's general implications for the behavioral treatments of autism. We believe that the model has implications far beyond autism; we discuss some of them in Section 5.

2 The Basic Framework

A child is exposed to a large variety of sensory data: visual, auditory, olfactory, and tactile. There is too much data for anyone to absorb; somehow a child must decide what to focus on, or treat as relevant, while ignoring the rest. Then the child must somehow represent the data that is viewed as relevant. Although the focus in this paper is autism, we believe that to get a good understanding of autism, it is important to understand how children decide what is relevant and how they represent data. While getting a deep understanding of how this is done may take decades, we already have enough of an understanding at this point to be able to sketch a workable model. In sketching the model, we try to be clear on what we take to be the status of various assumptions we make. We start by considering the

issue of representation, and consider later how the data might be "filtered" so that a child focuses on what he or she thinks is most relevant.

We assume that, after filtering, the incoming data stream consists of relatively short, discrete units. In the case of auditory data, these units could consists of phrases in some language, snippets of songs, the barking of a dog, and so on. In the case of visual data, the input could consist of sequences of visual "scenes".[2] It is not yet clear how the sensory inputs provided by a scene or utterance and the associations between them are represented by neural structures or activities in the brain. But they clearly must be represented somehow. We use a standard tool of discrete mathematics, and describe the representation by a *network*; that is, a collection of nodes and links. The nodes are labeled by the inputs and the links describe the associations between them.[3] We further assume that each node and link has associated with it a *weight* (which we take to be a real number between 0 and 1) that, intuitively, describes a combination of how frequently the data item labeling the node has been seen and the importance of the context in which it has been seen. Again, it is clear that there must be an analogue to weight. Any data that is learned will eventually be forgotten in the absence of further reinforcement. We can think of the weight as a measure of how long it will take to be forgotten. The weight is also correlated with the likelihood of retrieval. Our assumption that the weight of nodes increases with the number of observations of the data labeling the node and with the importance of the context in which the observation was made is consistet with recent views of memory mechanisms as being adaptive [Nairne et al. 2007].

A portion of a child's network is described in Figure 1. This network has a node labeled with the auditory input "Jon", with an associated weight of .8, and another node labeled with the auditory input "drinks milk", with a weight of .4. The node labeled "Jon" is joined by a link with a weight of .7 to a node labeled by a visual input, presumably a visual image of Jon, with a weight of .6.

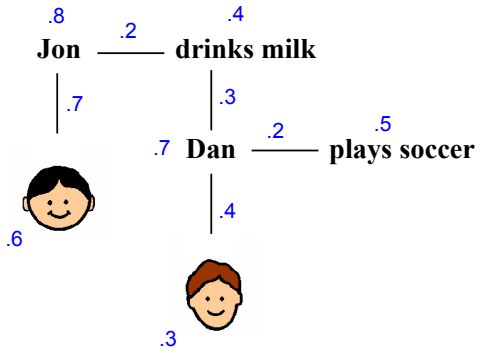

Figure 1: A portion of a child's network.

Determining exactly what physical constructs in the brain correspond to the nodes and links in our

[2]We are implicitly thinking of "continuous" visual data as having been discretized, like a movie, which consists of a sequence of static scenes.

[3]The use of such an abstract representation for inputs and associations resembles the use of concepts such as genes and alleles in population genetics, long before the discovery of DNA. A "gene" was initially an abstract term describing the presumed representation of a trait in the genome, where a trait was operationally defined (as color, size, shape, behavioral strategy, and so on).

network representation is beyond the scope of this paper. But discussions of this issue arise frequently in the literature. For example, temporal binding theory [Brock et al. 2002; Milner 1974] suggests that nouns and objects may be encoded by specific neuronal units, while their unique combinations in time and space may be coded by synchronous neural activity. If this theory is true, then, at least for inputs corresponding to words, the nodes in our model can be taken to be neuronal units and the links to be the synchronous activation of them. Similarly, if the neuronal clique theory [Lin et al. 2006] of encoding memory is correct, the neural cliques (network level coding units) are the nodes, while the links are the neuronal structures or temporal activities that assemble these cliques into hierarchies.

We are not the first to suggest using networks to represent data. Indeed, some features of our model resemble those of other approaches that involve data representation by networks; however, the details are quite different. Moreover, some critical elements in our model have no analogues in current frameworks in the literature. Thus, rather than using a particular pre-existing modeling approach such as artificial neural networks or cortical feature maps (see, for example, [Gustafsson 1997]), we use what is in some sense the simplest model needed to illustrate the basic principles of our approach. We discuss the relationship between our framework and others later in the paper.

2.1 Constructing the Network

To be useful, a representation must be well organized. We expect the network to be used for searching, goal-directed behavior, and problem solving (in ways that we make clear below). The link structure of a well-constructed network must make it easy to do all these things. As we shall see, the crux of our theory is that the network of autistic children is poorly constructed, with particular systemic problems that cause the symptoms of autism.

There are a number of obvious types of associations that we expect a well-constructed network to represent, including temporal co-occurrence of data items, spatial co-occurrence of data items, and similarity of data items. Keeping track of the fact that caterpillars often crawl on leaves is clearly important to a bird; for similar reasons spatial and temporal co-occurrence is relevant to hunters. Temporal co-occurrence is also critical for causality, while similarity is critical for generalization.

The network in Figure 1 illustrates these associations. A normal child who has heard a phrases like “Jon drinks milk”, “Jon”, and “milk” at various times might end up with such a network. The link between the nodes labeled “Jon” and “drinks milk” represents the fact that these phrases were heard at almost the same time.[4] The other links can be similarly explained.

There is a simple technique that ensures that the network will be constructed this way. Roughly speaking, we assume that the first time a child hears a phrase such as “Jon drinks milk”, a node labeled “Jon drinks milk” is added to the network. When later the phrase “Dan drinks milk is heard”, we assume that the child recognizes that both phrases share the common subphrase “drinks milk”. The label “Jon drinks milk” is then segmented to “Jon” and “drinks milk”, with a link (representing temporal co-occurrence) joining the nodes with these labels. A node labeled “Dan” is also added, with a link to the node labeled “drinks milk”.

Hearing the same phrase repeatedly increases its weight. The fact that the node labeled “Jon” has a weight of .8, while “drinks milk” has a weight of only .4 indicates that “Jon” has been heard more

[4] A slightly more sophisticated representation might use a *directed* network, where the links have arrows. In this case, the link might be directed from “Jon” to “drinks milk”, to represent the fact that “Jon” was heard just before “drinks milk”. We do not need to assume this level of sophistication for our results, although the brain might well keep track of this.

frequently (or in more significant circumstances—see below) than "drinks milk". We assume that the weights labeling nodes and the links that connect them decay over time unless they are refreshed by further examples/observations. We further assume that there is a *threshold* weight; once the weight of a node or link reaches this threshold, that node/link is fixed in memory, and, in the case of the node, its label is not further segmented.[5]

Decay and fixation are critical to our framework. Decay allows mistakes to be eliminated (we do not want children to remember random co-occurrences), while fixation preserves reliable data units and the links between them. They are clearly both adaptive; moreover, there is a great deal of empirical support for them. Work on memory shows that some data seems to be fixed in memory, and is remembered over long stretches of time. The fact that we do not forget an old telephone number suggests that the node representing that number is fixated. Similarly, the fact that the smell of a particular perfume may trigger a memory of a romantic evening suggests that the link between node representing the smell and the node representing the romantic evening is fixated (as are the nodes). On the other hand, memories do fade over time, showing that there must be physical construct that corresponds to weights decaying. It is clear that we need something like a threshold to avoid over-segmentation. For example, we do not want to segment "carpenter" into "car" and "penter".[6]

The idea that there is a critical window for segmentation, between when the weight of a node hits the fixation threshold and when it decays to 0, is perhaps the main novel idea in our representation. We can think of this as the critical window during which learning can occur. Because the segmentation of sequences takes place between decay and fixation, an inappropriate data distribution can result in poor segmentation of data sequences, which in turn can result in a network with an "inefficient" link structure, whose nodes are labeled by long data sequences. In particular, if one data sequence s is observed repeatedly, and data sequences that share some common subsequence with s are observed only rarely, then s may reach fixation before being segmented, and thus be represented by a single node. For example, if an autistic child hears "Don't touch the stove" repeatedly, while not hearing other phrases that involve "don't", "touch", or "stove", then "Don't touch the stove" may become fixated. A consequence of this is that, rather than using the word "stove" appropriately, the whole phrase will be uttered. We stress that the fact that the child does not hear "don't", "touch", or "stove" in other phrases does not necessarily mean that such phrases are never uttered in the child's presence. Rather, it means that, due to the child's filter, the child does not pay attention to these phrases. Thus, when we talk about "inappropriate data distribution" here, we simply mean the data distribution *after* the child has filtered out data he views as irrelevant is different from what the algorithms typically get; this includes insufficient data and data with an atypical distribution.

Our examples have considered segmentation of words. Similar examples involving segmentation of images result in links that represent spatial co-occurrence. Besides links representing temporal and spatial co-occurrence, we believe that a well-constructed network should also have links that represent similarity. The weights on a link that represents temporal or spatial co-occurrence can be thought of as

[5] Although, for convenience, we are viewing the threshold as sharp, it is perhaps more realistic to think that the higher the weight, the less like it is that the weight decays. This change has no significant impact on our approach.

[6] Segmentation is further complicated if there are a number of alternative ways in which data sequences can be segmented based on commonalities (a problem that increases when more data is available). The decision regarding how to segment in this case may be based on the relative weight of the alternatives (an example is given in Section 4) or, more generally, by finding the most appropriate segmentation for the context (as emphasized by Solan et al. [2003]). For example, the sequence "carpenter" may not be segmented to "car" and "penter" because it is accompanied by an image of a carpenter rather then by an image of a car. We discuss the problem of finding context in more detail in Section 3.

capturing the frequency and/or importance of the co-occurrence. (For example, even if the association of hissing with a snake does not occur frequently, the link might well have high weight because of the importance of this co-occurrence for survival [Nairne et al. 2007].) By way of contrast, the weight on a link that represents similarity captures the degree of similarity. (We are implicitly assuming that links are "tagged" with type, so that a link representing temporal co-occurrence can be distinguished from a link representing similarity.)

We do not consider here the details of how similarity is recognized. We assume that there is a process that can compare two data items for similarity, including similarity of subcomponents. While this is certainly a significant assumption, it is consistent with recent studies (see [Sloutsky 2003] and the references therein). Moreover, however it is done, we would argue that such a process *must* exist, since it is essential to identify different representations of the same data item. For example, a child may have seen the dog Fido in a number of contexts, and from a number of viewpoints (in the house, on the field, from the front, and from the rear), and have separate nodes for each of these instances of Fido. At some point, the child may realize that the dog labeled Fido seen from the front is the same as the dog labeled Fido seen from the rear, and coalesce the nodes labeled by these two images. Such coalescing plays a significant role in learning. It means that what the child has learned about Fido in different contexts can be combined.

Recognizing similarity is also critical for doing generalization. For example, suppose that a network includes a number of nodes representing people, such as "Jon", "Dan", "Shania", and "Louisa". These nodes are not similar enough to be coalesced, since they are associated with somewhat different visual images. On the other hand, they are similar in that they all have links to nodes like "rides a bicycle" (or just "rides" and "bicycle", if "rides a bicycle" has been segmented), "drinks milk", and many others. By observing the commonality, generalization is possible—these nodes can be grouped into a concept such as "child". Formally, what this means is that a new node labeled "child" is created, with links from the nodes labeled "Jon", "Dan", and so on to the node labeled "child". These links can be viewed as "is-a" links. Jon is a boy, boys are children, children are people, people are living things, and so on.[7] The subnetwork of a network consisting of the nodes constructed by the generalization process can be viewed as what is called an *is-a hierarchy*. The importance of is-a hierarchies has been frequently noted in the AI (Artificial Intelligence), psychology, and neural network literature [Brachman 1985; Fahlman 1979; McClelland and Rogers 2003]. We argue later that generalization is also critical for having what has been called "mentalizing abilities" in the literature [Baron-Cohen 2000], that is, the ability to attribute mental states to others.

For a network to be used effectively, we must assume that there are other processes that can effectively search and explore the network. For example, suppose that a child wants milk. We assume that the child then searches its network for a node associated with the feeling of drinking milk. Suppose that the child finds a node linked to the feeling of drinking milk. Such a node is linked to an image of milk and a phrase such as "want milk" (or to "want" and "milk", which are themselves linked). Once a relevant node is located, the links can be used to go from the desire for milk to a request for milk. In general, we assume that the exploration is done in a way that takes advantage of the weights, so that links with higher weight are explored before links with lower weight. Exploring a node or link in the process of the search may also have the effect of increasing its weight. Thus, after a number of searches

[7] Although we are assuming different types of links—links that represent temporal and spatial co-occurrence, links that represent similarity, and links that represent generalization—it seems plausible to us that the weights on all these links might have the same physical realization, in terms of the degree of activation. However, as we have said before, our framework is independent of the exact realization of the network.

for a particular item, the search itself will go faster.

In our discussion, we have assumed a number of processes, in particular, processes that segment data, recognize similarity, search a network, and so on. We have been deliberately vague here on how these processes work. While there is a reasonable consensus that these processes exist, there is far from a complete understanding of how they actually work; this is an area of active research. For example, there has been a great deal of work on how infants do word segmentation (see [Brent 1999; Jusczyk 1999] and the references therein), and a number of different models have been proposed. In addition, there has been a great deal of work on generalization in the context of neural nets (see [McClelland and Rogers 2003] and the references therein) and on searching in networks (see [Kleinberg 2001] and the references therein). Both bodies of work emphasize the importance of having links that reflect local structure and that model hierarchical structure (i.e., is-a hierarchies). Not only is there active work in this area, the work has led to theories that do not always agree. (For example, there are quite different theories on how word segmentation is done [Brent 1999].) Fortunately, for building a general theory of the type we are trying to build, the details of how these processes work do not matter. All that matters is that they exist.[8]

2.2 The Data-Acquisition Disorder and the Evidence For It

The claim that different data distributions are likely to result in different representations is hardly surprising or controversial. But this is not enough to explain autism. Not all "nonstandard" data distributions result in autistic behavior. Clearly, deaf or blind children get data distributions that differ from those of normal children, and do not typically develop autism. What is important for our claims is that the data distributions encountered by autistic children (that is, the distribution of the data after filtering the vast amount of data that the child is exposed to) have the following two properties.

- There is relatively little data about people and their behaviors (including social interactions).[9]
- Some phrases (or non-verbal data sequences) are encountered repeatedly in "significant" contexts (one which are given high weight), and their constituents are encountered rarely in other contexts. For example, a phrase such as "Don't touch the stove" may be encountered relatively frequently, while "don't", "touch", and "stove" may be encountered rarely in contexts other than "Don't touch the stove!".

Note that the first property leads to networks with relatively few nodes related to people and their interactions; moreover, the nodes involving people and their interactions that are in the network are likely to have relatively low weight. The second property leads to networks whose nodes are labeled by relatively long data sequences, with inefficient link structure. To understand why, note that because we assume that segmentation is due to similarity between parts of data sequences, if a long phrase like "Don't touch the stove!" is heard repeatedly without hearing "touch" and "stove" in other contexts, "Don't touch the stove!" will eventually approach fixation and be stored as one data unit, without being

[8] We remark that in our computer simulation in Section 4 we do make further assumptions about the segmentation process for definiteness but, as we point out there, other segmentation processes should give qualitatively similar results.

[9] There may well be other domains for which an autistic child may have relatively little data, but all that matters for our analysis is that they have relatively little data about people.

segmented.[10] Similarly, repeated attention to a particular dog image in a favorite children's book may result in storing this image as one data unit before the observation of the features of dogs in other images allows for the segmentation of the image into some shared shared visual components (eyes, ears, tail, and so on).

Having a network whose nodes are labeled with many long phrases, rather than a network whose nodes are labeled with short (typically single-word) phrases linked together is the effect of the second property. (Similar comments hold for images as well as phrases, of course.) Note that we would not expect the data distribution of blind or deaf children to have these two properties, even though they are likely to be quite different from those of normal children. What is important is not just that data distributions differ, but *how* they differ.

How reasonable is it to expect autistic children to suffer from such a data-acquisition disorder? To examine this issue, we first consider the evidence for such data-acquisition disorders more generally.

Extensive research has been conducted on the mechanism of filial and sexual imprinting in animal behavior [Bateson 1966; Bateson 1979; Hauber et al. 2001; ten Cate and Vos 1999]. These studies have shown that, in many young animals, during sensitive developmental periods, a set of innate preferences for particular stimuli (such as particular fragments of a bird song) help guide the search for the most suitable model of a parent, a future potential mate, or a tutor for song learning [Bateson 1979; Hauber et al. 2001; Soha and Marler 2000]. These innate preferences are refined by experience through the imprinting process, ensuring that further social attachment and learning will develop appropriately. In other words, future learning is based on acquiring relevant data by following, imitating, and learning from the correct models. Although no natural data-acquisition disorder has been reported in this respect (which is not surprising, considering that such a disorder makes survival quite unlikely), a simple version of such a disorder can easily be induced in captivity. By merely replacing the correct model with an inappropriate model, the young of many birds and mammals can be imprinted on other species, on their human caregivers, or even on their empty cage [Bateson 1966; Bateson 1979]. Such animals are usually incapable of acquiring typical social behavior and are unlikely to survive in the wild. The potential risk of misimprinting is considered on a regular basis when raising young animals to be released in the wild, and has been implicated in shaping the evolution of animal behavior [Lotem 1993; Rodriguez-Girones and Lotem 1999]. The reason, we believe, is that misimprinting leads to interpreting the "wrong" data as relevant, and thus results in a highly skewed distribution of data relative to a normally-imprinted animal.

Autism in humans is clearly a much more complex disorder than that resulting from artificially-induced social misimprinting in animals. Yet humans also seem to have a set of innate preferences that direct their social orientation and early cognitive development. It is not important for our purposes whether these preferences are completely innate or are extended by pre- and post-natal experience. It is clear, however, that human babies not only seek a warm human touch, but also express a great interest in human voices and faces, and even the direction of adults' eye gaze [Haith et al. 1977; DeCasper and Fifer 1980; Emery 2000].

All these initial preferences have been found to be impaired in autistic children (at least to some degree). Autistic children typically display a marked lack of interest in other people relative to other children [Baron-Cohen and Bolton 1993; Howlin 1998]. More specifically, they are less likely to follow eye gaze [Baron-Cohen et al. 1997; Klin et al. 2003], to be interested in human faces [Hobson et al.

[10] Even if it is segmented into "Don't touch", "the" and "stove!", there will be links with high weights between these three segments. This will result in them being treated as a "tight cluster", from which it is hard to make connections. (We return to this issue at the beginning of Section 3.)

1988; Pierce et al. 2001], or to attend to human voices [Klin 91]. It has been suggested that this, in turn, leads to problems in joint attention and imitation [Carpenter and Tomasello 2000; Frith 1989; Heyes 2001; Mundy 1995]. In addition, many autistic children apparently derive little positive feedback or enjoyment from watching, touching, being with, and being touched or hugged by other people [Frith 1989; Howlin 1998]. All these impairments provide evidence for the first property that we assumed, namely, that autistic children acquire relatively less data about people. Autism researchers are certainly aware of these problems in social orientation. Some have even proposed that deficits in social orientation and joint attention might be the primary cause for autism [Carpenter and Tomasello 2000; Heyes 2001; Insel et al. 1999; Mundy 1995]. However, to date, there has been no explanation of how these basic deficits can cause the more complex deficits and cognitive abnormalities observed in autistic children.

As we mentioned earlier, data-acquisition problems in autism are not related only to social data. Occasionally, autistic children either display extremes of sensitivity to stimuli such as touches, sounds, or smells (being either very sensitive to them or displaying very little sensitivity to them), although there is no evidence that the problem is in the sensory mechanism itself. This has commonly been viewed in the literature as a sensory-processing problem that has to do with the way the brain responds to input [Anzalone and Williamson 2000; Frith 1989]. Some of these problems have been viewed as problems that emerge from enhanced perceptual functioning [Mottron et al. 2006]. The causal relationship between sensory or perceptual abnormalities and impaired social interest in autism is not yet clear. It is possible that the latter is caused by the former, or that both are caused by a more basic neurological problems. Whatever the cause, the result is an atypical data distribution.

Finally, children along the autistic spectrum have a tendency to exhibit interest in peculiar and unusual subjects (such as a strong interest in vacuum cleaners), and to insist on acquiring and memorizing detailed data about these subjects [Baron-Cohen and Bolton 1993; Frith 1989; Tager-Flusberg et al. 2001]. Again, this results in an atypical distribution of data.

These problems in data acquisition have a number of effects. The lack of interest in other people results in observing and hearing less relevant data per unit of time, since much of our auditory input comes from listening to other people. Recall that one effect of this in our model is that nodes and links are more likely to decay before becoming fixated, which will result in fewer opportunities for successful learning. The lack of joint attention may also result in auditory input being associated incorrectly with visual input [Carpenter and Tomasello 2000]. A higher-level problem is that an autistic child may not care if listeners are following his own eye gaze or center of attention. This further reduces the probability of obtaining any answers, or correct answers, which in turns reduces the number of correct links. Moreover, it reduces the motivation to communicate, thus resulting in even less data. The lack of interest in people can also result in certain (intuitively inappropriate) parts of a scene or a sentence being viewed as "interesting", while other parts are ignored. For example the scene "Jon drinks milk" may be taken as a scene of "drinking milk", with Jon being ignored, or a movement of a man's fingers may cause an autistic child to be interested in the hands but not in the person [Baron-Cohen and Bolton 1993]. This will result in inappropriate scenes being linked to certain utterances.

Despite their general reduced interest in people, autistic children may still be interested in close or familiar people or in things that occur under particularly rewarding contexts (such as contexts that provide food or a high level of excitement). But this too may result in a skewed distribution that leads to inappropriate fixation. For example, a lack of general interest in human speech, combined with a strong interest in a mother and a new baby sister may lead to an autistic child hearing "touch" and "baby" only in the context "Don't touch the baby", when the phrase is related to something that he wants to and is

about to do. As a result, the phrase "Don't touch the baby" will get fixated, rather than being segmented as a result of hearing "baby" and "touch" in other contexts. It is commonly observed in autistic children that they use long phrases like "don't touch the baby" when they see a baby. Often these phrases are lines in songs about animals or objects, or phrases taken from television shows or children's books. This phenomenon, where a child "echoes" a phrase such as "Do you want milk?" when she asks for milk, is known as *echolalia* or *echolalic speech*, and is typical among autistic children [Frith 1989; Howlin 1998]. In fact, to a large extent our model was inspired by the phenomenon of echolalic speech, which intuitively seems like a problem of data segmentation. Interestingly, the literature on word segmentation makes no references to echolalic speech in autism or to any similar disorder of word segmentation. Moreover, in a review of research on word segmentation, Brent [1999] comments that very little reliable and reproducible data about children's segmentation errors is available. It seems that researchers on word segmentation have not noted the possible connection between autism and segmentation problems, which is a central theme in our framework.

It is much more difficult to obtain similar evidence for segmentation problems in nonverbal data. But given the evidence for such problems in speech data, we make the leap of assuming that similar problems occur in other contexts. As we show in Section 3, making such a leap, together with all the other assumptions we have made, leads to a model with a great deal of explanatory power.

2.3 Where Does the Inappropriate Data Distribution Come From?

Why might an autistic child have an inappropriate data distribution? As we suggested in Section 2.2, this certainly might be an outcome of well-known problems of autistic children with social orientation and joint attention. But is there a deeper causal explanation of these problems in data acquisition?

There is a great deal of data available to a child. Most of it is ignored, to prevent sensory overload. How does a child decide what to pay attention to? This is a critical issue since it effectively determines the child's distribution of data. (Recall that when we talk about a child's data distribution, we are referring not to the distribution of data *available* to the child, but the distribution of the data that the child actually pays attention to.) We now propose a mechanism where the network itself can determine what the child pays attention to. While this mechanism is not critical to our argument that autism is a result of problems in the distribution of data, we view the fact that we can use the network itself to determine relevance, rather than needing to assume some external mechanism that somehow "knows" what is relevant, as a significant advantage of our approach.

The idea is that an individual will pay attention to a data sequence to the extent that it matches items already in the network. The better the match, the more attention will be paid to the item. This view is consistent with our view of the network as a way of organizing the individual's view of the world. Not only does the network keep track of associations, it keeps track of degree of importance. It is clearly adaptive for an individual to pay attention to important items. For example, we would expect that, for an animal, the image of a hiding place, the sense of distasteful food item, or the sound of a predator's footfall are all important. These are items that should have high weights in a "good" network for the animal.

But the network structure allows us to go beyond just looking for matches to nodes with high weight. For one thing, parts of a data sequence that is viewed as "interesting" that do not match pre-existing nodes themselves become labels of new nodes, and can be used to "attract" more new data from the environment. Thus, we get associative learning. For example, when networks are built this way, a bird

can learn to focus on a certain type of leaf after seeing a tasty caterpillar on such a leaf a few times.

A second advantage of thinking in terms of the network structure is that the network can be used to deal with conflicting matches. There may be a great deal of data that matches to some degree with nodes in the network. We expect individuals to focus on those data sequences that match nodes with higher weight. Again, this is adaptive: we should focus more on more important data rather than on less important data. We observe this phenomenon all the time. A child already very interested in cars will naturally focus on cars. For similar reasons, we would also expect an autistic child interested in vacuum cleaners to focus on them. Of course, the effect of focusing on data sequences that already have high weight in the network will be to increase their weight. It also has the effect of bringing into the network data items related to data items in the network.

The assumption that individuals represent their information using a network suggests that there must be some initial network-like structure. In biological terms, it can perhaps be thought of as a genetically preconfigured neural structure that represents some data items and the links between them, to which external input can be compared. The positive feedback cycle means that quite small differences in the initial network can lead to quite different networks in older children. The existence of such an innate initial network is very much in keeping with the presumed "innate templates" observed in animal behavior [Bateson 1966; Bateson 1979]. As we noted earlier, several studies suggest that both animals and humans have a set of innate preferences that guide their search for the most suitable model of a parent, a social mate, or a tutor for learning and imitation. It is also known that animals have innate responses to some key stimuli related to recognizing food, predators, and social partners [Shettleworth 1998].

We can think of the initial network as a "starter", around which the network develops. The nodes of this network will essentially act as "attractors" for matching data from the environment, and will focus the attention of the individual on certain data items and not others. By analogy, the network grows around the initial nodes like a crystal that crystallizes around initial starters. Having inappropriate starters can result in attention being focused inappropriately, which in turn results in getting an inappropriate data distribution. Under this viewpoint, the data-acquisition problems of autistic children can be traced to an inappropriate initial network or some "damage" to the network during the first years of life. The causal structure of our framework is summarized by the scheme described in Figure 2.

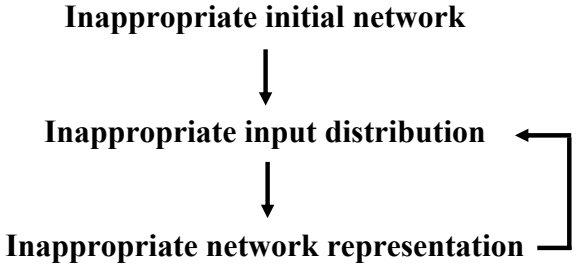

Figure 2: The causal structure of the framework.

Of course, to the extent that this is correct, differences in the initial network clearly would have an impact well beyond autism. For example, they can help explain sex differences of the kind studied by

Baron-Cohen et al. [2002]. It seems possible that small genetic differences between males and females can result in differences in their initial network. For example, suppose that females' initial networks typically have them focus more than males on data about things like human facial expressions, eye gaze, body language, voice tunes, and other features that are related to people's social behavior and emotional states, while males' initial networks result in males focusing more on data about people's physical attributes, such as how they run, jump, hunt and fight, and about their physical environment (e.g., features of their landscape or objects in the environment that can be used and manipulated). These initial differences, when amplified by the positive feedback loop, can result in females having networks that have much more data about other people, and so (as discussed in Section 3.1) being better at mentalizing and empathizing with others, while males have networks that lead to their being better at manipulating the environment and what Baron-Cohen et al. [2003] call "systemizing", that is, constructing explanatory theories of how the world works, and constructing and controlling systems.

Baron-Cohen and his colleagues [2002, 2003] have used this empathizing-systemizing dichotomy as a way of explaining autism. Our model is consistent with the view that high-functioning autism and Asperger syndrome are like an extreme male condition in terms of failure to acquire human data and developing an empathizing network. One possible explanation of this similarity in terms of our model is in terms of similarities in the initial network of an autistic child and a child with extreme male condition. However, it might also depend in part on the fact that it may require more data to learn how to attribute mental states to others (see the discussion in Section 3.1 below). Systemizing of simple systems can be done with relatively little data, so that even a severely autistic child is likely to systemize something (e.g., he can probably find patterns in his collection of toys). Thus, even if there is no problem in the initial network of an autistic child, he may end up exhibiting symptoms of an extreme male condition due to an inappropriate data distribution.

We remark that this view of the empathizing-systemizing spectrum as arising from differences in an initial network is compatible with Badcock and Crespi's [2006] genetic imprinting theory of autism. Roughly speaking, Badcock and Crespi view autism as the outcome of a "conflict" between genes coming from the father and mother's side. To the extent that the father's genes are expressed, we get more systemizing; to the extent that the mother's genes are expressed, we get empathizing. In our framework, the expression of the genes is reflected in the initial network.

2.4 Summary and Related Work

We have discussed how a child constructs a network, with weights on nodes and links. Clearly, the more (accurate) data a child has, the more useful the network constructed is like to be. For example, in a "good" network, a word is linked to a visual image if that word describes the image. Of course, for a child's network to have the right links, the child must be looking at the image that matches the word(s). In this respect, our approach is consistent with the view that joint attention and social orientation are critical for word learning [Bloom 2001].

More relevant data is also likely to lead to smaller sequences, since with more data, there will be more opportunity to detect similarities, so more sequences will be divided into subsequences before reaching the weight threshold. However, it is not just the quantity of data that matters, but the rate at which it arrives and its distribution over time. Given the decay process, if data arrives too slowly, links and nodes will disappear. On the other hand, having certain data sequences arrive too frequently is not necessarily good either. For example, if a data sequence is frequently reinforced before similar

sequences are observed, it will be fixated before it has a chance to be broken up. Thus, there is a critical window of time during which sequences can be compared and broken into subsequences before they are forgotten or fixate.

To summarize, decay and fixation combined with "normal" observations should result in "natural" units (single words like "dog", "cat", and "Jon" and corresponding visual images) becoming fixated, and errors decaying, while inappropriate distributions of data sequences can lead to somewhat arbitrary data sequences becoming fixated, even if all the processes are working correctly. In addition, as we suggested above, a network constructed from appropriate data should help the search process (both because there will be "better" is-a hierarchies and because there will be more links between relevant related items). Moreover, a well-constructed network helps provide context. That is, if a node like "cat" is accessed, then we would expect there to be links with high weights from "cat" to other nodes like "milk", "mouse", and "Scout" (the name of a specific cat) as well as the appropriate visual images. There would also be links with perhaps lower weights to "tree", "climb", and "bird" and their related visual images. On the other hand, we would expect very weak links (or no links at all) to "kangaroo", "car", or to actions like "read" or "drive". Clearly, the outcome is likely to be quite different if a network is constructed from inappropriate data. As we argued briefly above (and illustrate in much more detail in the following sections), inappropriate data distributions can result in segmentation problems that lead to poor network representation. The poor network representation, in turn, can produce the three core deficits of autism.

Although there are clearly many details that need to be fleshed out, the model has a number of appealing features. For one thing, although it is relatively simple, it is the first to provide a principled developmental mechanism able to explain many of the symptoms observed with autism. Moreover, it suggests why intensive early-intervention programs—the method currently considered most effective for dealing with autism—are effective: they help autistic children to acquire the intensive flow of data that typical children acquire spontaneously.

To the extent that the model we present is accurate, then we would almost certainly expect there to be a disorder corresponding to the breakdown of the processes that we have sketched; we happen to give the name "autism" to that disorder. The similarity between the "autism" that our model predicts and the symptoms of autism that are actually observed provides support for the model. Note that our model is compatible with the view that autism has a genetic basis—indeed, it suggests what mechanisms such a genetic predisposition might affect.

The idea that primary deficits related to joint attention and social orientation may be the cause for autism is not new (see [Mundy 1995; Bushwick 2001] for attempts to explain autism along somewhat similar lines). However, in our model, we define these primary deficits in terms of data-acquisition problems, and provide a clearer developmental mechanism that explains how such primary problems can generate the core deficits of autism and other related symptoms. Note that our approach is different from a static neuropsychological deficit approach, in which the brain is characterized in terms of a normal brain with some parts or "modules" impaired [Shallice 1988; Temple 1997]. Instead, it resembles recent approaches that view developmental disorders as deviations in developmental trajectories that lead to brain abnormalities [Oliver et al. 2000; Johnson 2000; Thomas and Karmiloff-Smith 2002] (see also [Pierce et al. 2001; Miller et al. 2003] for supporting evidence).

Our model has features that appear in many other frameworks, such as neural nets, statistical learning, chunking, and case-based reasoning. It would in some ways be convenient if our model could be embedded in one of these frameworks. However, we believe that attempting to do so at this stage would

be premature for two reasons. First, some critical elements in our model, such as the active selection of data input, and the dynamic interaction between the window of decay and fixation and the segmentation process, simply have no analogues in other current frameworks. Second, we have attempted to build a high-level theory that uses the minimum number of principles necessary to explain the observed phenomena. Embedding our approach into another would require additional, perhaps unnecessary, assumptions. We do cite some of these other frameworks at various points in the paper, but given the vast quantity of literature, our comparison is rather cursory.

The possibility that an inappropriate network may explain some cognitive characteristics of autism has been suggested by several neural net models [Cohen 1994; Gustafsson 1997; McClelland 2000; Gustafsson and Paplinski 2004]. Our model provides more of a causal explanation of how an inappropriate network may develop. We also use it to explain a far broader set of phenomena associated with autism. In fact, we claim that much of what is known about autism can be explained in terms of our model; this point is discussed in great detail in Section 3. We recognize that some of the mechanisms that have been suggested in the neural-net models mentioned above may be consistent with, or incorporated in, our framework. For example, excessive lateral inhibition [Gustafsson 1997] may be a possible cause of the abnormalities of the initial network in our framework. There are also some similarities between the idea that learning with familiarity preference can generate inappropriate networks [Gustafsson and Paplinski 2004] and our idea that an individual whose network has nodes labeled by long data sequences is likely to focus on objects with similar "large" labels, and therefore, his network will not develop in a typical way.

While our main goal in constructing the basic model is to use it to explain autism, the model itself, if correct, should also be consistent with the large body of evidence in the field of language acquisition and cognitive development. A comprehensive analysis of this issue is beyond the scope of the paper. Nevertheless, our model does in fact seem to be consistent with many recent findings. For example, it should be clear that our model is compatible with statistical learning. As we have observed, the weights on the nodes and links can be viewed as representing the statistical frequency of observing data sequences and the frequency of their association with each other. Evidence for statistical learning by 8-month-old infants suggests that they are capable of using transition probabilities for word segmentation [Saffran et al. 1996; Aslin et al. 1998]. Although in our model we do not specify an algorithm that monitors transition probabilities, the algorithm that breaks up data sequences based on shared components does much the same work. It breaks up data sequences into smaller subsequences if the transition probability between the subsequences is relatively low, and tends to keep sequences together if their transition probability is relatively high. This is because a low transition probability implies a higher probability of acquiring data with matching segments before the sequence as a whole reaches fixation. For example, the transition probability between "Jon" and "drinks" in "Jon-drinks" is likely to be considerably lower than the transition probability between "cof" and "fee" in a natural word like "coffee", given the likely transitions from "Jon" to "eats" and from "Dan" to "drinks". Thus, there is a good chance that the sequence "Jon-drinks" will be broken by the sequence "Jon-eats" before it gets fixated. On the other hand, "coffee" is likely to reach fixation before there is an attempt to break it up by a word like "toffee" or "coffin". This example also illustrates the importance of the time window between decay and fixation. Thus, our model would predict exactly the same results as those reported by Saffran et al. [1996].

More sophisticated forms of statistical learning can also be supported by our model. For example, a recent study by Gomez [2002] suggests that 18-month-old infants can learn dependencies between the first and third words in a sentence even when the second word is unpredictable. Such learning can

occur in our model as a result of having links between all data units that appeared in the same data sequence before segmentation. The fact that dependencies between adjacent words can be learned at a much younger age then long-distance dependencies is not predicted by our basic model. However, it can be incorporated into the model by assuming that early in infancy weights decay more rapidly, causing infants to keep links only within very short data sequences, which are likely to be heard more often. Having such a mechanism can help in building a "better" network by ensuring that, early in development, small data units, like words, reach fixation before they can be "locked" into commonly-used phrases.

Our model does not have a hard-wired innate "universal grammar" module in the sense of Chomsky [Chomsky 1965; Pinker 1994]. Nevertheless, we believe that it is consistent with (and, indeed, provides a framework for) statistical, data-driven language acquisition, in the spirit of recent work in the area (see [Elman 1990; Elman 1999; Pullum 1996; Solan et al. 2005] and recent evidence reviewed by Tomasello [2000]).

Not only is our model compatible with statistical learning, it is also compatible with reinforcement learning. All we need to assume is that some nodes can be associated with positive or negative connotation (perhaps by being linked to appropriate centers in the brain, so that they can act as reinforcers or motivators). We do not go into further details here.

Finally, we observe that the basic model we have described has many similarities with work on *chunking* [Gobet et al. 2001], as well as with approaches that use recurrent neural nets to find words and syntactic structures in time (e.g., [Elman 1990]). Our model is different from models of chunking mainly because it assumes a mechanism that builds the network by breaking up data sequences, and by having a dynamic system of weights that allows modification of the network, and creates the critical window between decay and fixation. In contrast to neural nets, we assume that nodes represent real-world phenomena directly. (Of course, patterns of nodes can be viewed as representing more complex phenomena. For example, seeing Dan's blue T-shirt will stimulate a number of nodes.) Nevertheless, assuming that nodes are directly related to real-world phenomena suggests certain algorithms for breaking data sequences and finding commonalities between sequences that would not be as appropriate (and, we believe, more difficult to implement) in the more standard neural network setting.

3 The Effects of Data-Acquisition Problems

As we suggested in the introduction, given the importance of getting appropriate data, we would expect there to be some disorder associated with a defect in the data-acquisition stage. If it were not autism, it would have been something else. Moreover, as we have argued, some of the symptoms of autism do indeed suggest that autistic individuals suffer from data-acquisition problems. In this section we show how our model can explain the main cognitive deficits associated with autism. For simplicity, the discussion here is couched in terms of verbal utterances, but the same comments apply to visual scenes as well.

We first briefly summarize the effect of the observations. Given what we know about the data distribution of autistic children, we would expect that the verbal utterances associated with nodes in the network of autistic children contain relatively more phrases that have not been broken up into their constituents. Moreover, even when phrases are broken up, we would expect more "clustering" than in normal children, with strong links (i.e., links with high weight) between nodes in a cluster and weak

links between nodes outside a cluster. For example, words or phrases that appear repeatedly in a song or story that the child particularly likes and thus pays attention to will be strongly linked to each other, but not to other words or phrases to which they are likely to have links in normal children.Because longer phrases are associated with nodes, and longer phrases are less likely to be similar to each other than individual words are, fewer nodes will be identified as similar and be unified through the coalescence process. As a result, we would expect that, in order to represent the same amount of data, the network of an autistic child would typically have more nodes than that of a normal child, and these nodes are also likely to be labeled by longer, more idiosyncractic data sequences. Moreover, while the overall number of links in the network of an autistic child might be larger than the number of links in the network of a normal child (because the autistic network has more nodes), the number of links to each individual node is likely to be smaller, since fewer nodes are viewed as similar.

Since the network of an autistic child is composed of large components (either nodes labeled by long data sequences or large clusters) that "lock" smaller words or images inside them, it represents the world in a fragmented and inflexible way. A natural unit like "dog" will appear in a number of clusters (for example, a cluster associated with a song that mentions a dog, another cluster with a story involving a dog, and so on). There will be strong links from an occurrence of "dog" in a cluster to all the other nodes in that structure, but only weak links to nodes associated with dogs in other clusters. An autistic individual might store the word "dog" in many different clusters, much as a normal individual might store the syllable "car" within different words like "carpenter", "carnival", or "carnivore" in different clusters. Moreover, since the links emanating from the various "dog" nodes will lead to quite different nodes, these nodes will not be generalized to a notion of "dog". The strong links within a cluster offer immediate access to what the dog has done in that cluster, but hampers the search for "dog" in other clusters that may suggest what else the dog could do. In such a network, search, retrieval, or simulation within a cluster is easy, but search, retrieval or recombination of components from different clusters is difficult. As a result, generalization in such a network will be more difficult, local processing will be easier than global processing, and rigidity and perseveration will be more likely than flexibility and creativity. In other words, this network can be expected to produce what some workers view as an autistic cognitive style [Frith 1989; Happé 1999; Russell 1997].

The fragmented nature of the network of an autistic child may also explain why autism looks like a state of retarded concept formation [Snyder et al. 2004]. The process described by Snyder et al. of forming concepts and metaconcepts from collections of sensory details is somewhat analogous to that of generalization and building is-a hierarchies. If sensory details are locked in their original sequences, the only concepts that can be formed are those that describe precisely the way these details were first observed. (See also Section 3.4, where we relate this issue to the phenomenon of autistic savants.)

We now discuss how these properties of the network can lead to the cognitive deficits associated with the three leading theories of autism. We deal with each of them in turn.

3.1 Theory of Mind

One central core deficit in autism is impaired theory of mind, sometimes known as *mind-blindness* [Baron-Cohen 2000; Frith 2001]. This core deficit refers to the inability of autistic individuals to attribute mental states to others and to predict their thoughts and feelings. Recent neurological research associates particular brain areas with the ability to mind read or to mentalize, and suggests that neuroanatomical abnormalities in these parts of the brain may be responsible for mentalizing failure in

autism [Frith 2001; Gallagher and Frith 2003]. The possibility that mentalizing is carried out by specialized modules in certain parts of the brain is reasonable. However, thinking in terms of specialized brain modules does not completely explain exactly what goes wrong in autism.

The cognitive mechanism of mentalizing is not clear. It is quite evident that both animals and humans have basic mechanisms that allow them to make some predictions about future behaviors of other individuals, including that of rivals or predators [Hasson 1991; Shettleworth 1998; Zahavi and Zahavi 1997]. This basic ability, which is also viewed as "mind reading" by behavioral ecologists (e.g., [Krebs and Dawkins 1984]), may be facilitated by an innate tendency to follow the eye gaze of others [Emery 2000] or to imitate [Heyes 2001], or it may be based on experience [Emery and Clayton 2001]. However, discussions of theory of mind usually consider a higher cognitive ability that is more sophisticated then predicting someone's behavior based on some cues. Humans (and possibly some primates) are postulated to have cognitive representations of the "mental states" of others; that is, a "theory of their mind" [Baron-Cohen 2000; Leslie 1987]. Researchers in the field argue that the representation of mental states must somehow be decoupled from the representation of the real world, and should be marked "in quotes" to signify that they represent someone's mental states rather than real physical facts [Frith 2001; Leslie 1987]. The crucial problem, left unexplained, is how a neural mechanism can possibly acquire the necessary information for decoupling representations and identifying them correctly as "mental" versus "physical".

For example, it has been shown that by 18 months of age, a child who watches his mother using a banana as a telephone will know that she is pretending [Frith 2001, p. 970]. Even if we assume the existence of a highly sophisticated neural system that has evolved specifically for these tasks, there is still a problem of explaining how such a system can tell that in this case the banana is intended to represent a telephone, and not the other way around. Since a telephone is a recent innovation in terms of human evolution, the recognition of a mother's intentions regarding the banana could not be innate. This is different from cases such as an animal's escape response to a predator's eye gaze [Zahavi and Zahavi 1997], where the recognition of the predator's intentions towards its prey could, at least in theory, be innate. Thus, to explain the more advanced mentalizing abilities of children, we must propose a mechanism that interacts with the environment and learns to identify a wide range of possible mental states and attitudes. We now consider in more detail how the network can support the attribution of mental states to others, how it can be used to understand pretend play and false beliefs, and why all these abilities are likely to be impaired in autism as a result of having an abnormal network.

Attributing mental states to others: Attributing mental states to others is a necessary first step in developing a theory of mind. Only then can these mental states be monitored and aggregated to create a more complex mental map of others' attitudes and thoughts. Although we can never feel the pain, the emotions, the thoughts, or the desires of other people, there is accumulating evidence that when we observe other people in a particular context we can simulate our own feelings in the same context and attribute them to the observed individuals [Ramnani and Miall 2004; Singer et al. 2004]. However, there are good reasons to believe that this ability is switched on only after making the appropriate generalization: classifying the observed individuals as people like oneself. Unless such a generalization is in effect, there is nothing preventing us from attributing our feelings to a washing machine or a pocket calculator. We do not expect mentalization to be immediate; it depends on the appropriate generalization.

For example, if Jon has learned that Dan has many of the same links to actions and attributes that he

himself has, he can generalize that when Dan falls on the floor, he will also feel pain, or that when Dan looks at the ball, he also sees the ball. More precisely, in Jon's network, there is a node representing the general concept of a boy, which has links to Jon himself, and to other boys like Dan, and to attributes or actions that boys have, or can do. Based on Jon's own experience with mental states he can now attribute such mental states to all boys.

In order to generalize appropriately, the appropriate units must be available. That is, it is impossible to have a concept of "boy" (that can feel pain and see a ball) unless the network has nodes for individual boys like "Dan" and "Jon". Moreover, in order to recognize that "Jon" and "Dan" belong to the same category (i.e., to establish an "is-a" link from each of them to the node representing the general concept of a boy), they must share at least a certain number of features in common. Thus, appropriate data acquisition is critical for both the initial identification of the individual units like "Jon" and "Dan" (otherwise the data sequences will decay or reach fixation before being broken up) and for the accumulation of sufficient data about them.

This explains why autistic individuals have difficulty with the basic ability of attributing mental states to others. In the most extreme case, young autistic children that have little interest in other people might still be at a stage where they do not even classify an individual like Dan as a boy, and therefore would not associate certain observations of Dan with the appropriate mental state that they themselves may experience in a similar context. In terms of our network model, the node representing "Dan" may not be linked to a node that represents the concept "boy", either because there was too little data to create such a node in the first place, or because not enough data was collected specifically about Dan to establish a link from the node "Dan" to the node "boy" (if such a node already exists in the network). In either case, the result is that there will be no links that can eventually lead from "Dan" to the mental states experienced by the autistic individual in a similar context.

In the more benign cases, autistic individuals may be able to categorize an individual like Dan as a boy, but might still have difficulty in identifying his mental states. One reason is that the fragmented nature of the autistic network might result in there being several nodes that represent the concept "boy", but each may be linked to a limited set of mental states. Dan, for example, may be matched with a particular node labeled "boy" (or, perhaps better, "boy1", to distinguish it from other nodes associated with boy-like features) that is linked to sport activities and with mental states like "happiness after wining a soccer game" or "desire to kick a ball". However, "boy1" may not be linked to a mental state such as "likes cars", which may in fact be linked to another node "boy2" representing the generalization of different boys.

A second reason for difficulties in attributing mental states to others is that even when Dan, for example, is classified correctly as a boy, and even if an autistic individual recognizes that he himself is also a "boy", the node representing the concept "boy" in the autistic network may not be linked to many potential mental states. This is because autistic individuals do not make many observations of the behavior of other boys that results in such links. For example, Jon, an autistic individual, may classify Dan as a boy because he looks like a boy, talks like a boy, and is very similar to a group of other individuals who were called "boy" by his mother. Yet, because Jon was never really interested in the behavior of these boys he is not aware, for example, that boys usually scream or protest when they fail to get what they want—which is exactly what he, Jon, does in the same context. Thus, while Jon associates the situation that produced the scream with the feeling of "anger" for himself, he will not be able to identify the similarities in behavior and context, and thus infer that perhaps other boys can feel "anger" too. That is, in Jon's network, there will be links between "Jon" and "anger" and between "Jon"

and “boy”, but not between “boy” and “anger”.

With sufficient experience, even an autistic individual may be able to create links between “boys” and “anger”, but these links are likely to be much weaker than for normal individuals. The probability of following such weaker links is lower. So, in practice, that means that autistic individuals are less likely to attribute mental state spontaneously or easily, although they might be able to do so upon request and with some effort.

Understanding pretend play: There are a number of issues involved with pretending. The ability to pretend in the sense of deceiving someone else is more related to the ability to understand false beliefs, which will be discussed in the next subsection. Here we focus on the ability to understand pretend or imaginative play.

Consider the example mentioned earlier of a child who watches his mother using a banana as a telephone. Based on Leslie’s [1987] approach, Frith [2001] suggested that in order to understand that his mother is pretending, the child has to understand her mental attitude towards the banana (i.e., that she wants to pretend that the banana is a telephone). To achieve this, the child needs to decouple the secondary representation of the banana as a telephone from its primary representation as an edible fruit, and to remember somehow that the secondary representation is limited to the context of his mother’s pretend play. We now show how our model can explain how this can be done, and why it is likely to be difficult for autistic children.

In the network of a nonautistic child, there is already at least one cluster of nodes representing previous observations of people talking over the phone. It includes nodes representing the sounds and the visual scenes of a person talking and holding a telephone with links to a node representing the telephone as an object. This cluster is probably also linked to nodes representing words and phrases like “telephone”, “mom is on the phone”, or “let’s talk over the phone”. The child also has in his network clusters that include a node representing “banana”. Most likely, the cluster that includes “banana” has very few links to the cluster associated with using a telephone. At some point, the child encounters for the first time a situation in which his mother uses a banana as a telephone. It would appear that, based on the information in his network, there is not much he can do except to treat this new scenario as a rare but real combination in which mom is using a banana to talk over the phone. In fact, we would argue that if the mother was pretending very well, and using the banana exactly as she uses a telephone (without giving special attention or hints to the child), the child may well be fooled into believing that bananas can occasionally be used as telephones, at least by his mother.

However, when engaging in pretend play with a child, many additional cues are provided. For example, a mother may be sitting with her child on a carpet, playing with other toys (most of which are themselves pretend objects, such as animals or people), and suddenly start using the banana as a telephone while still keeping her typical play intonation. Moreover, we claim that these cues can often be matched and linked to cues in other nodes in the network that are associated with some general features of typical pretend play. To justify this claim, we must explain how typical features of pretend play could become labels of nodes in the network in the first place. We do not want to assume that they were there initially, nor that there is a genetic program that defines “pretend”. Rather, we suggest that this concept is developed in the network just like any other concept: by acquiring data and building appropriate links that eventually establish two parallel subnetworks, one for real scenarios and one for pretend scenarios.

For example, a child typically acquires a great deal of data about people driving and using real cars,

and about similar scenarios in which people, and especially children, use toy cars. The two parallel behaviors share many similarities, so many of their nodes will be linked, or even shared. Yet, each set of behaviors also has some unique features. For example, toy cars are smaller, are used mostly in the play room, and are associated with some unique words like "toy". In fact, quite early in life the child may hear words or phrases such as "real" versus "not real" or "pretend" or "it is just a toy", and will learn to associate them with the right sets of objects and behaviors. Toy cars are obviously just one example. There are dolls, puppets, little houses, pretend food, and even a pretend telephone. So, by the time that a child is faced with his mother using a banana as a telephone (at 18 months in Frith's [2001] experiments), he already has a fairly rich representation in his network of real versus pretend scenarios, with links to the appropriate words describing each set.

Now we can return to the child's ability to understand his mother's pretend play where she is using a banana as a telephone. While the mother's behavior is clearly similar to using a telephone, it is missing the match with a real telephone. Moreover, it has several contextual and stylistic features that make it similar to typical play. As we explained above, the child can probably find in his network many cases in which play behavior paralleled a real scenario, and where mimicking reality was part of the game. Based on the structural similarity of these pretend situations to the current situation, the child is likely to infer that his mother's behavior is an instance of pretend play. (This inference would, of course, be particularly easy if the child had already played with a toy telephone before.)

Note that making such an inference requires an "is-a" link to be established, and to conclude that a new situation "is-a" pretend play, without ever being told so explicitly. The generalization mechanism is exactly the one discussed earlier in the context of attributing mental states to others. At some point, when the situation is sufficiently similar to other instances of pretend play, the child concludes that this is also an instance of pretend play. One implication of our model is that recognizing the situation as pretend play comes first; based on this recognition, the child can then conclude that his mother is pretending, or that she is treating the banana as a telephone. This is quite different from the approach according to which the child somehow reads his mother's mind and understands that she is pretending, and thus that the whole situation is an instance of pretend play. In our view, because a child cannot really read his mother's thoughts, the only feasible way to infer her mental attitudes is by reading the situation and the context, and fitting them to similar situations.

We can now see why autistic individuals are likely to have problems in understanding pretend play. First, as in the case of attributing mental states to others, inferring that a novel situation is an instance of pretend play depends critically on having sufficient data to conclude that the novel situation is "similar enough" to other known forms of pretend play. Autistic children who fail to acquire sufficient data about the basic forms of pretend play will not be able to generalize the concept of pretend play to more sophisticated games. Second, the units that are being compared in the case of pretend play are large clusters of nodes that represent complex situations. The task of finding similarities between such clusters will be especially difficult for autistic children because, in their network, such observations are likely to be stored in different clusters of nodes, with little linkage between these different clusters. In other words, understanding that a situation is pretend play requires putting it in the context of pretend play. As we discuss extensively in Section 3.2, putting things in context is very difficult with the fragmented network of autistic individuals.

Understanding false beliefs and developing a theory of mind: The deeper meaning of having a theory of mind goes beyond the ability of attributing mental states to others, or the ability to distinguish

reality and pretense. By the age of four, nonautistic children appear to be able to monitor the mental states of others and to create a mental map, or a theory about what other people see, know, believe, or even feel [Baron-Cohen 2000]. The most notable evidence for this skill is the child's ability to understand false beliefs. This is probably the major stepping stone to more sophisticated behaviors involving the ability to tell lies and to manipulate the knowledge or the mental states of others. One of the special requirements for achieving a theory of mind is the ability to separate between one's own knowledge about the world and the knowledge attributed to someone else [Birch and Bloom 2001]. By analogy, this is like having separate "folders" in a computer in which we store and update data about different peoples' attributes, knowledge, and mental states. Having a theory of mind has therefore been viewed as an advanced cognitive skill that may require unique modalities [Frith 2001]. While we agree that developing a theory of mind is a highly nontrivial process, we do not believe that it should require any special mechanisms beyond those considered already in our model.

In our model, a "folder" is nothing more than a node with many outgoing links. For example, a folder for "Dan" is just a node labeled "Dan" with links to nodes that, intuitively, give information about Dan. Once a child reaches the stage where he is attributing mental states to others, there is nothing to prevent another child's mental state from being included in the folder associated with that child. For example, if smiling faces are associated with the mental state of being happy, when Jon meets a new child in the playground with a smiling face, the mental state of being happy would be linked to this new child, just like his dark hair and green shirt.

One key use of these folders is to construct (portions of) the networks of others. For example, if Jon sees Dan staring at a ball in a box, and Jon knows that when he (Jon) looks at a ball he adds a link from the ball to the box in his (Jon's) network, he may infer that Dan adds a link from the ball to the box in his (Dan's) network. Thus, Jon starts to build a rudimentary representation of Dan's network in the Dan folder. A more sophisticated problem is that of representing that Dan knows about things Jon does not know about. For example, if Dan looks in a box and Jon does not know what is in the box, Jon has to represent (in his representation of Dan's network), that Dan knows what is in the box, although Jon does not. Although it is beyond the scope of this paper to go into the details of how this could be done, we remark that the standard representations of knowledge in the philosophy and computer science communities using graphical structures usually called *Kripke structures* (see [Fagin et al. 1995]) can easily be embedded in the networks we have constructed.

With this view in mind, we can now consider the task of understanding false beliefs, and the reasons for its impairment in autistic individuals. For concreteness, we focus our discussion on the most commonly-cited false-belief test, which is the *Sally-Anne task* [Baron-Cohen et al. 1986]. In this task, a child is presented with a scenario involving two girls, Sally and Anne, and a ball, a basket, and a box. The scenario is presented using puppets or illustrations that describe the following sequence of events. Sally puts her ball in the basket, and then goes away. After Sally has left, Anne moves the ball to the box. When Sally returns, the child is asked where Sally will look for the ball. By providing the correct answer ("in the basket"), the child shows that he is able to identify Sally's false belief that the ball is still in the basket.

Answering a question such as "Where is the ball?" essentially amounts to using the network to perform goal-directed behavior. However, to answer a question like "Where will Sally look for the ball?" requires the child to perform the goal-directed behavior in his representation of Sally's network. Note that the realization that this is required is also a generalization: the child generalizes from the realization that he searches his own network to answer such questions to the fact that all children use

their networks to answer such questions. Of course, to make such a generalization, the child must already have a relatively sophisticated representation of other children's networks. Since, as we have already argued, autistic children are likely to have a rather poor representation of other children's mental states, it is not surprising that they come by this realization relatively late, if at all. But even if the child does realize that it is necessary to search (his representation of) Sally's network to answer the question, this is likely to be a more difficult question for the autistic child than for a normal child, again because an autistic child is likely to have a poor representation of what Sally knows. The fact that autistic children are typically less interested in faces and people will make them less likely to store the relevant information about Sally. Finally, even if they have the relevant information, they might not have made the generalization from the fact that they search for an object the last place they saw it to the fact that others also search for objects the last place they saw them.

Many features of the network have to work together in order to support a theory of mind. It is necessary to identify individuals, accumulate sufficient data about individuals in order to be able to generalize and to infer their mental states, and then have enough data to draw inferences about things like reality vs. pretense and typical behaviors (like looking for something where it was last known to be). Given all these requirements, it is perhaps not surprising that some deficits in theory of mind can characterize even the most highly-functioning autistic individuals. Impaired theory of mind can thus be viewed not as a primary problem, but as an unavoidable symptom of a data-acquisition disorder.

3.2 Weak Central Coherence

The term "weak central coherence" has been attributed to a broad set of experimental findings, showing how the performance of autistic and non-autistic individuals differs, mainly in the perceptual domain (see the review by Happé and Frith [2006]. Typically it refers to the superior ability of autistic individuals to process details, and their relative impairment in extracting global meaning [Frith 1989; Happé 1999]. The exact mechanisms involved, and the relationships between enhanced local processing and reduced global processing are still unclear [Happé and Frith 2006; Mottron et al. 2006; Plaisted et al. 2006]. Weak central coherence has also been used to describe the impaired ability to put new data in context [Frith 1989; Happé 1999]. As we now show, this impaired ability is in fact predicted by our theory. With a few further assumptions, we can also show that the differences between autistic and non-autistic individuals in extracting details and getting the big picture follow from this impairment.

The structure of the "abnormal" network of autistic children suggested by our model makes it difficult to put new data in context. To see why, suppose that an individual tries to understand a word like "dog". He presumably hears the word and then scans his network to find the most similar nodes, that is, nodes associated with the word "dog" or a similar-sounding word. (Recall that we have assumed that such similarity-based searches can be carried out.) For a young autistic child, there may not be any exact match for "dog", since the word has not yet formed a separate unit. Rather, the word will be a part of a larger sequence (such as "See the dog"), which will make the similarity search more difficult. For an older child, on the other hand, the problem is that there may be an embarrassment of riches. Through the years, such a child might have enough opportunities to form the word "dog" as a separate unit (either by hearing it separately, or by breaking up a few larger sequences). In addition, "dog" may already have been locked in a number of other clusters (far more than for a normal child). Hence, there may be many exact matches for "dog", each in its own cluster, that have not coalesced into a single node. We assume that one of these matches will be chosen, but it will not necessarily be the most appropriate one, given the context. Since we expect that "understanding" will come by following links coming from the node

chosen, it is not as likely that an autistic child will have as good an understanding of "dog" in context as a normal child. In particular, an autistic child will not have as rich a set of connections to "dog" as a normal child.

It may well be that if one match does not succeed, an autistic child will try to find another match. But this process will take time, especially if there are many possible matches. The experimental evidence also suggests that, at least in some cases, autistic children stop trying to find matches before finding the "right" match [Frith 1989; Jolliffe and Baron-Cohen 2001a]. It is also possible that, in some autistic individuals, the repeated failure to put things into context attenuates the motivation to look for matching data in old memory, thus creating a "lazy habit" to process information locally rather than globally. This will weaken central coherence even further.

Interestingly, there is also evidence that autistic individuals often try to put things in context and sometimes may be successful in doing this [Brian and Bryson 1996; Lopez and Leekam 2003; Mottron et al. 1999; Ozonoff et al. 1994; Ropar and Mitchell 1999]. This apparent inconsistency is not surprising if we do not view weak central coherence as a general cognitive style, but rather as the result of trying to use context in an abnormal network. The failures to use context may then reflect abnormalities in the part of the network needed for the relevant task. Different individuals may have different abnormalities, depending on their particular developmental history.[11]

In the following, we review some of the evidence for weak central coherence in the literature and show how it can be explained by our model.

- Hermelin and O'Connor [1967] observed that, in contrast to control subjects, autistic individuals do not derive benefit from meaning in memory tests. In particular, they do not recall words from a sentence better than words from a random list of words. This can be explained using our model as follows. We assume that when hearing a sentence such as "The cat climbed on the car" in a memory test, a node labeled "sentence to memorize" is created, with pointers (links) to the first word ("The"), the second word ("cat"), and so on. Note that nodes associated with these words are presumably already in the network. These words are not intended to be novel words; it is only their co-occurrence in a particular combination of words that is new and needs to be memorized. However, if the co-occurrence itself is not novel, then a nonautistic child will presumably already have links between many of these nodes; for example, there is likely to already be a link between "cat" and "climbed" and between "climbed" and "on the car" (or between "climbed" and "on", and between "on" and "the car"). Having these links will help in the recall task for a nonautistic child. After recalling "The cat", it will be easier for a nonautistic child to remember the rest of the phrase, just by following pre-existing links. These will lead to a set of candidate words to complete the sentence, from which the best match with recent memory can be selected.

  By way of contrast, an autistic child is less likely to have such pre-existing links (or, if they exist, they are likely to be weaker links). Even if he has a link between "cat" and "climbed" and a link between "climbed" and "on the car", these may be two different instances of "climbed" that have not coalesced, so it will not be possible to follow the link from "cat" to "climbed" to "on the car". Instead of recalling each word independently based only on its recent link (pointer) to the node "sentence to memorize", the pre-existing links can support additional recalling channels. On the

[11] Recall that the process of segmentation itself may be assisted by context. Thus, we may have a positive feedback cycle here: basic segmentation problems can lead to a network representation that makes it difficult to use context, which may in turn further increase segmentation problems.

other hand, for a string of random words, autistic and nonautistic children are likely to be on an equal footing when it comes to inter-word links.

So far we have considered only inter-word links. However, the recall of meaningful sentences may also be helped by links between words and images and between those images. People frequently visualize words and verbal descriptions; this visualization helps in memorization and recall. Autistic individuals may also be impaired in this respect, due to the poor link structure between nodes that represent words and nodes that represent visual images. While a nonautistic child can easily visualize the sentence "The cat climbed on the car", an autistic child may be unable to do so. Even if the word "cat" is associated with a particular image of a cat, this particular image may be locked within a specific picture of, say, a cat sitting on the carpet, with no further links to other cats climbing, and no other images of the act of climbing on a car.

- Frith and Snowling [1983] and Snowling and Frith [1986] showed that autistic individuals fail to use context to resolve the pronunciation of homographs. For example, they have more difficulty than normal individuals with the pronunciation of "tear" in "there was a big tear in her eye" and "there was a big tear in her dress". Typically, autistic readers respond with the most frequent pronunciation for a homograph, regardless of context. A simple explanation of this is that autistic individuals typically do not try to use context while reading [Frith 1989; Happé 1999]. As we mentioned earlier, such a "lazy habit" may be explained by our model as a secondary result of repeated failures to put things into context in general. However, there may also be a direct explanation. The fact that autistic individuals select the most frequent pronunciation suggests that they try to use existing links in the network (i.e., to use context), but that the links available to them may be different than those in a typical individual's network.

  In terms of our model, we would expect that the node corresponding to (the image of) the written word "tear" is linked to two possible pronunciations. The node corresponding to the pronunciation heard more frequently has, by definition, greater weight. When a nonautistic individual reads the sentence "there was a big tear in her dress", this frequency effect is outweighed by additional link between the word "dress" and the appropriate (but less common) pronunciation of "tear" (in a dress). In autistic individuals, on the other hand, the link between "dress" and the second pronunciation of "tear" may be weak or not immediately accessible. Even if both "dress" and "tear" are present somewhere in the network of the autistic individual, they might be represented by nodes that have not coalesced, each still locked within a different cluster. Hence, although reading the word "dress" may provide access to the visual image of a dress, and to other nodes linked to a particular instance of a dress (like "you look very pretty in this dress"), it may not be linked to the particular instance of "dress" that is linked to the the appropriate pronunciation of "tear". Without such a link, all that the autistic individual can do is to follow the stronger link from the written "tear" to its more common pronunciation.

  Interestingly, additional aspects of Frith and Snowling's work may also be explained by this view of "competition" among links. For example, they show that when autistic individuals fill in a blank in a sentence, they are able to use context for choosing a word from the correct syntactic category, but usually fail to select the most appropriate word in terms of meaning (despite having no problems in assigning semantic meaning to words). However, their performance improves when their set of choices is restricted. To explain these findings in terms of our model, consider for example the task of filling in the blank in the following sentence: "He looked in the _____ in the riverbank" (used by Frith and Snowling [1983]). The algorithm that is used to search for the

appropriate word in the network is likely to be based on searching the network for a certain amount of time, and then selecting the word with the most "satisfying" combination of links to all other units in the sentence. The degree to which a given set of nodes is most "satisfying" should clearly be some increasing function of the number of links between nodes in the set and their weights. A nonautistic individual may easily find many links leading from a node labeled "looked-in-the" to nodes like "box" or "basket", but perhaps only one or two of them also have pre-existing links leading to "riverbank". These might be words such as "cracks" or "cave" that are linked to "riverbank" (perhaps due to a cave having been seen in a riverbank on a TV program). It is quite possible that the word "holes", which is the original missing word in this sentence (which was taken from a story about beavers), will not be found in the search and another word will be selected.

Now consider an autistic individual who searches his network in a similar way, and puts in the same amount of effort as a nonautistic individual. His chances of finding what we generally consider a context-appropriate word are much lower. Even if he has heard many of the relevant words, including many instances of "riverbank", there might be several instances of "riverbank" in his abnormal network that have not been coalesced. Hence, one "riverbank" he finds may be locked in a cluster like "The boat approached the riverbank very slowly", while another may be locked in a cluster like "The children went swimming by the riverbank", and so on. Based on this search, among all words with links coming from "looked-in-the", the word "boat" may be the best choice, because it clearly also has a link to "riverbank". The search is still likely to lead to syntactically-appropriate words (as indeed observed by Frith and Snowling). This is because the strongest combination of links coming in this case from "looked" are likely to be to a syntactically-appropriate word like "boat" rather than to words like "approached" or "slowly". Past experience with a data sequence like "looked in the boat" is much more probable than with a sequence like "looked in the approached" or "looked in the slowly". (Note that the knowledge of grammar in our model is also data-driven.)

Finally, we have to explain how restricted choice can help autistic children to find the context-appropriate word. Consider again the sentence above, this time with the restricted choice offered to the children: "He looked in the holes/drawers/books in the riverbank" [Frith and Snowling 1983]. In this case, the competition between the strongest combinations of links is limited to the three options. This helps in two ways: First, searching backward from the possible solution may give access to the more relevant (albeit weaker) links much faster. Second, in the abnormal network of autistic individuals, there may be alternative combinations of links that are stronger than the correct one; the restricted choice helps eliminate them. For example, for an autistic person, the first easily accessible instance of "riverbank" found in the limited search may be the one linked to the word "boat", while another instance of "riverbank" that is in fact linked to the correct solution "holes" will not be accessed due to its lower weight in memory. (This problem does not exist in a normal network, where all nodes labeled "riverbank" are likely have been coalesced.) The restricted choice forces the autistic individual to disregard the first accessible solution and to search only among combinations that include the three solutions. This may actually prolong the search (as indeed reported by Frith and Snowling [1983, Table 7]), but helps an autistic child to perform at above-chance level.

- Jolliffe and Baron-Cohen [2001b] report that adults with high-functioning autism or Asperger syndrome show poor performance (relative to controls) in their ability to integrate a few visual

objects (drawings) into one coherent scene, or to identify the meaning of a scene, and to spot the odd object in this scene. Most of the tested individuals were eventually able to perform most of the tasks, but only after more time and with much greater effort than the controls. Jolliffe and Baron-Cohen conclude that the greater effort needed for autistic subjects results in their tending not to process in context unless they are instructed to do so or they consciously decide to do so. However, they do not explain why it is so difficult for autistic individuals to process in context.

To explain these results with our model, consider first the experiment in which the subjects had to integrate four out of five line drawings into a single coherent scene, and to identify the incongruent drawing among those five (the one not needed for the coherent scene). More specifically, we focus on an example in which there are five drawings: a boy, a ball, a dog, a tin can, and a fence. The "correct" answer involved forming a scene of a boy throwing a ball to knock cans off a fence (leaving "dog" as the incongruent object). This problem requires searching the network to find the strongest combination of pre-existing links that can integrate four of the five objects. For a nonautistic individual, the strongest combination is indeed likely to represent the most reasonable scene, because the weight of the links is based on past co-occurrence of sequences. However, making use of these links in the abnormal network of autistic individuals is likely to be much more difficult. As we observed in Section 2.2, for an autistic individual, nodes such as "dog" are likely to appear repeatedly in the network, each time locked in a cluster with only a few links to other nodes. This will make finding the best combination much harder. For example, to compute the strength of the combination of "dog" and "fence" will require looking at all instances of "dog" and "fence" in the network. In addition, since the network of an autistic individual is likely to be based on an abnormal distribution of data sequences, the strongest combination of links may not reflect correctly the most likely co-occurrence of pictures in reality.

All the other experimental results reported by Jolliffe and Baron-Cohen [2001b] can be explained by our model in a similar way.

- In a related paper, Jolliffe and Baron-Cohen [2001a] provide evidence regarding some even more basic aspects of visual processing. They show that autistic individuals have difficulty integrating fragments of a composed object and identifying its holistic meaning. For example, given separate drawings of a hinge pin and two rectangular plates, they have difficulty integrating them to form a hinge [Jolliffe and Baron-Cohen 2001a, p. 215]. However, their ability to identify an object based on the visual image of a broken piece of that object (e.g., a tip of a key) is unimpaired.

  To explain these results with our model, consider first how visual images are stored in the network of nonautistic individuals. Most newly-observed images are actually composed of components that have been seen before. As a result, according to our model, they will be broken by the data-processing algorithms and stored as clusters of nodes with the appropriate links between them. (The idea that image representation is based on feature integration is not unique to our model; see, for example, [Treisman and Gelade 1980; Cave and Franzel 1989].)

  By analogy with verbal data, in many cases, a visual image will be stored in the network not as a single word, but as a sentence or even a paragraph. For example, a bus may be stored as wheels (or even circles) linked to a box that is also linked to a row of small squares (the windows), and a "train" may be stored as a chain of such "buses". This type of representation allows components to be reused, making for greater efficiency. In autistic individuals, however, the situation is somewhat different. Each node in their network is likely to represent a longer data

sequence.

Now consider Jolliffe and Baron-Cohen's example of identifying a hinge. If in a nonautistic individual the image of this hinge is stored as a cluster of nodes consisting of a "pin" linked to two "rectangular plates", then there is a good chance that the fragments in the drawing will be matched with nodes labeled by similar images in the network. Integrating the fragments (which are now represented by nodes) into a single coherent object can then be done by finding the strongest combination of links that connects all three visual nodes together, and by naming this object using additional links to a common node at the verbal level.

For this process to work, all the relevant instances of "rounded pins" or "rectangular plates" in the network must be either linked or coalesced. This important requirement is exactly what is not typically achieved in the fragmented network of autistic individuals. In the extreme case, the composed image (i.e., the hinge) may be stored as a single node, without any links between its potential fragments. Even if the image of the hinge actually consists of several components, the fragmented nature of the network makes it unlikely that the first match of the drawing fragments will be with the particular pin and plates that form the required cluster of a hinge. Due to the limited degree of coalescing in the network of an autistic individual, there are likely to be multiple representation of "pin" and "plate".

This problem should disappear if a fragment is sufficiently rare and unique that only one node in the network can be found as a reasonable match. In this case, upon finding the node, it can be linked only to the single possible target object. This explains Jolliffe and Baron-Cohen's observations that autistic individuals were unimpaired when the fragments used in the experiments were no longer separate components, but rather a unique part of an object, such as the broken tip of a key.

- One of the most frequently-quoted examples related to weak central coherence is the superior performance of autistic individuals on the embedded-figure test [Shah and Frith 1983; Happé 1999]. For example, autistic individuals may be better than controls at finding a hidden shape, like a triangle, within a drawing of a baby carriage (see [Happé 1999, Figure 3]). To explain this enigma, it has been suggested that autistic individuals are not seduced by the gestalt of the main object, and find the parts as salient as the whole [Shah and Frith 1983]. However, no specific mechanism has been proposed to explain how autistic individuals can see the parts so easily. Based on our model, we suggest that autistic and nonautistic individuals may differ in their performance on this task because they are likely to differ in the way they segment the salient object into parts.

  We expect that, in the network of nonautistic individuals, the baby carriage is stored as a group of nodes (visual fragments) connected by links. These fragments correspond to naturally-occurring objects such as the sun roof of the baby carriage and the poles. Although the embedded figure of the triangle is present in the final image, it is not in any of the fragments. Indeed, each side of the triangle is likely to have strong links to a (different) natural object. To identify the embedded triangle, a nonautistic individual must thus actively ignore the links leading to these natural fragments. In autistic individuals, on the other hand, the salient object may be stored in the network either as a single node, or as a tightly-knit cluster of nodes (with relatively few or very weak links between its nodes and similar nodes in other clusters). The absence of links that usually put the components in context now serves to limit distractions while searching for the triangle. This

makes the task much easier for autistic individuals.

- Children with autism are better than normal children at visual search tasks, and seem to possess enhanced discrimination ability. For example, O'Riordan and Plaisted [2001] asked subjects to perform the following four tasks:

  – searching for a red X hidden among green X and red C distractors;

  – searching for a red X hidden among pink X and red C distractors;

  – searching for a red F hidden among green F and red E distractors;

  – searching for a red F hidden among pink F and red E distractors.

  Normal children did significantly better on the first task than the other three, and found the last task hardest (intuitively, since pink is close to red and E is close to F). Autistic children did better than normal children on all four tasks. Although they also found the last task the hardest, their performance did not vary much between the four tasks. O'Riordan and Plaisted did not provide an explanation for this phenomenon. Again, our model can help here, provided we make some assumptions about the discrimination process.

  As we mentioned earlier, we assume that there is a process that can compare data sequences and identify shared components based on their similarity. As a special case, such a process should be able to compare a data sequence with a previously-stored sequence to see if they are essentially the same. It seems reasonable to assume that this check for sameness is easy if the nodes are (almost) identical or if they are very different; it is most difficult if they are reasonably similar but not almost identical. For example, imagine comparing two nodes associated with colors. Since a particular observation of colors is subject to some variation (depending on lighting conditions, etc.), as well as some degree of random error in the perceptual process, there is a greater chance of misclassifying two colors as the same if they are similar than if they are far apart. A simple approach to dealing with this is to use two thresholds: classify two colors as the same if the observation indicates that the difference between them is smaller than the lower threshold, classify them as different if the observation indicates that the difference is larger than the upper threshold, and repeat the observation (perhaps several times) if the difference is between the two thresholds to try to get an observation that allows for a more definite classification (for example, by averaging repeated observations). This two-threshold scheme already explains why, in general, it takes longer to distinguish red from pink than to distinguish red from green. Pink is likely to fall between the two thresholds, thus activating repeated observations to minimize errors.

  Recall that, according to our model, autistic individuals are likely to store an image like "red F" using a single node labeled "red F", while a nonautistic individual is more likely to have two linked nodes labeled "red" and "F". Thus, when comparing "red F" to "pink F", autistic children just have to compare one pair of nodes, while nonautistic children have to compare two pairs of nodes ("red" vs. "pink" and "F" vs. "F"). Thus, the amount of time required to make comparisons will already be significantly less than for nonautistic children on all tasks.

  Note that this explanation may be applicable to other examples of enhanced perceptual functioning in autism (see the review by Mottron et al. [2006]). If a visual image that is represented by a combination of data units in the network of non-autistic individuals is represented by a single larger unit in autistic individuals, when queried about this data, autistic children will be able to

respond faster because they will save on integration time and have the data more immediately accessible. Thus, according to our theory, there is no difference between the search and discrimination algorithms of autistic and nonautistic individuals, nor in their speed of operation. It is merely the way the data is stored in the network that gives autistic individuals superior perceptual abilities in some cases. At the same time, this atypical representation can impair the ability of autistic individuals to find context and to extract global meaning.

3.3 Executive Dysfunction

There is a large body of evidence suggesting that one of the core cognitive deficits of autism is executive dysfunction [Pennington and Ozonoff 1996; Russell 1997]. Since the scope of executive functions is broad and loosely defined (see [Hill 2004; Ozonoff 1997]), we focus on how our model can explain only the most central and relatively agreed-upon symptom of executive dysfunction, namely, the fact that autistic subjects persevere with incorrect responses in tasks that require a cognitive shift, as in the case of the Wisconsin Card Sorting Test (WCST).

In this particular test, as described by Ozonoff [1997], subjects are presented with four "key" cards, which differ in terms of the color, shape, and number of objects depicted on the card. For example, one card may have five blue stars; another may have three red circles. Cards in a deck are presented to the subject one by one, and the subject is asked to choose the key card that best matches the card turned over. The experimenter has in mind a specific rule ("match by color", or "match by number of objects"), which is not revealed to the subject. However, the examiner does tell the subject whether his choice is correct (according to the examiner's rule). Once the subject has made ten correct matches, the rule is changed, without notice or comment from the examiner. As a result, matches made according to the earlier rule now receive negative feedback. The primary dependent variable in this test is the number of trials in which the subject continues sorting according to the previously-correct rule, despite negative feedback. This number tends to be significantly higher for autistic individuals, testifying to their poor ability to make a cognitive shift [Ozonoff 1997; Pennington and Ozonoff 1996].

To explain this result using our model, we have to explain how an individual's network affects his ability to make the required cognitive shift in the WCST. We assume that a nonautistic subject has learned the concept of a game or test in which sorting cards or objects according to certain rules gave a verbal or a physical reward. This concept is thus a node in the network, to which are linked clusters of nodes representing instances of such games that have been played before. In particular, there are links to games played with cards or building blocks, that in turn are linked to nodes that represent concepts like colors, shapes, and numbers. (We are making the important assumption here that nonautistic children have a node representing "color", with links to specific instances of colors, and to past experiences in which objects were grouped based on their color.) The upshot of this is that shared components between games can be easily accessed. With such a network, a nonautistic subject can both recognize sorting rules (such as as "sort by color") easily and more easily manipulate concepts like color and shape to match the rule(s) used by the experimenter.

In the network of an autistic individual, the situation may be quite different. The atypical distribution of data is likely to cause particular games or tests to reach fixation in memory before they are split up into separate components. Note that this does not mean that autistic individuals are not good in deriving rules in general. In fact, it has been claimed that they are quite good at doing this [Baron-Cohen et al. 2003]. We would expect them to be able to quickly derive a specific rule from a single game that they

have focused on. In this case, an autistic child would have strong and exclusive associations between the games and its rules.

However, given the differences between the networks of autistic and nonautistic individuals, we would expect autistic individuals to have more difficulty in generating possible rules, even during the first stage of the test. For example, while many different games might involve sorting objects by a certain feature, such as size or color, the notion of sorting by color might be locked within the context of sorting blocks, while cards are remembered only as being sorted by suit. Thus, just as with weak central coherence, the probability of matching the new test context with a particular sorting rule stored in past memories is lower. The difficulty in generating new rules may in part account for why autistic subjects that manage to figure out the rule being used by the experimenter have so much difficulty in shifting to a new sorting rule.

There may be additional issues as well. When the examiner changes the sorting rule after ten successful trials, the rule previously used suddenly results in negative reinforcement. We would expect this negative feedback to lead the subject to search for alternative rules in the network. However, the simple rigid rule of shifting to the closest alternative behavior following the first failure is not generally adaptive (especially if it is the first failure after ten successive hits). Occasional failures are common in many natural foraging tasks and are tolerated at some level without causing a shift in behavior [Kacelnik and Bateson 1997]. We expect that the mechanism responsible for shifting from one behavioral strategy to another is based on a more quantitative balance between the expected payoffs of shifting and persisting. Roughly speaking, after "enough" negative feedback, an alternate rule would seem better than the currently-used rule. But if alternatives are hard to find, the greater cost of switching to an alternative for an autistic child will make them will make the currently-used rule seem better for a longer period of time. In other words, even if they have a fully functional mechanism for generating cognitive shifts, autistic subjects are likely to persevere with incorrect responses much longer than nonautistic subjects.

3.4 Other Symptoms

There are several other symptoms of autism that are not obviously consequences of the core deficits discussed above. Perhaps most notable are a tendency for repetitive behavior and adherence to routines [Baron-Cohen and Bolton 1993; Tager-Flusberg et al. 2001], and a collection of higher-level deficits such as a lack of creativity and problems in understanding abstract thinking and humor [Frith 1989; Howlin 1998]. There have been connections made between humor and difficulties with abstract thinking on the one hand, and the core deficits on the other [Emerich et al. 2003]. We briefly sketch how our model can explain some of these symptoms.

- Recall that the nature of the fragmented network of autistic individuals makes it easy to follow links within clusters, but not across clusters. Creativity is often viewed as the ability to make unexpected connections. To make such connections, it is seems necessary to have many "cross-cluster" links, which are precisely what autistic individuals lack.

- The ease in following just a few links may also explain some repetitive behavior: the behavior is repeated because there are no other available links to follow. Our model does not provide a direct explanation for the more simple repetitive behaviors like hand flapping or head shaking. It is possible that some of these behaviors have direct neurological causes. On the other hand, the presence of such behaviors in isolated or abused children [O'Connor and Rutter 2000] as well

as in captive animals [Garner et al. 2003] suggests that some of these behaviors may also be secondary outcomes of the disorder.

- Without going into the more complex issue of what humor is and why it exists, it is quite clear that understanding most types of humor depends critically on understanding context and being able to appreciate certain connections. Humor is often based on things that are either out of context, or fit into the context in a very unconventional way. Hence, autistic individuals, who have problems in understanding context, are likely to have difficulties in understanding humor. In addition, many of the more complex humorous situations require the ability to mentalize, which is also impaired in autism.

- Abstract thinking depends to a great extent on the ability to generalize and to create "is-a" hierarchies. We have explained in some detail why these abilities are impaired in autistic individuals and cause problems in developing a theory of mind. These points apply equally well to any type of abstract thinking. Without the ability to generalize, and to infer that things might exist even when they are not sensed directly, abstract thinking is impossible.

- Yet another "symptom" commonly associated with autism is that of the autistic savant [Hermelin 2001; Snyder and Mitchell 1999; Snyder et al. 2004]. The ability to remember in detail a long sequence of music, pictures, text or numbers can be explained to some extent by our model, in a manner much in the spirit of our explanation above of enhanced perceptual functioning. When a child hears a new data sequence, he tries to find a match to (part of) that data sequence in his network. If a match is found, the data sequence is found, the sequence is broken up, weights are changed, and the link structure of the network is modified. This process typically reduces the probability that a data sequence will be retrieved exactly in its original form. (Indeed, a teacher often takes it to be a sign of understanding that a student does not parrot back what was said earlier as an answer.) An autistic child is far less likely to find a match than a normal child, because his nodes are more likely to be labeled by long data sequences. Thus, if he stores the new data sequence in his network, he must store it as an entire unit. Of course, to store the data sequence at all requires that the child be interested in that data sequence. It may well be that most children are not interested in the effort required to store such long sequences. The reason that not every autistic child may show "idiot savant" style of abilities is that not all of them may have the initial interest in acquiring the type of data that allows them to exhibit such abilities. We thus agree with Mottron et al. [2006] that savant skills may develop through a brain-behavior cycle that requires a combination of special interests and enhanced perceptual abilities; our model provides an explanation for why perceptual ability is enhanced in autistic individuals.

- Finally, we remark that the fact that the network of an autistic child may be fragmented and have more nodes may also help explain some recent neuroanatomical findings that show that children with autism have larger brains than do comparison groups [Courchesne et al. 2004; Wallace and Treffert 2004], and that such brains exhibit local overconnectivity and long-range underconnectivity [Belmonte et al. 2004; Courchesne and Pierce 2005]. Some researchers have suggested that the enlarged brain and poor connectivity may be the causes of autism [Belmonte et al. 2004; Cohen 1994; Courchesne and Pierce 2005; Happé 1999]. Our model suggests an alternative explanation for these findings that may be equally plausible. The larger brain may be a plastic developmental response to the demand for extra memory capacity. The autistic brain continues to accumulate

long data sequences without being able to compress them based on similarity. This idea is consistent with the fact that brain size of autistic individuals is normal at birth, and becomes larger than average only during the first few years of life [Aylward et al. 2002; Courchesne et al. 2001; Courchesne et al. 2004; Wallace and Treffert 2004]. Similarly, according to our model, the local overconnectivity and long-range underconnectivity may not be the primary causes of autism, but rather a reflection of the fragmented network that develops in autistic individuals as a result of data acquisition problems.

4 A Computer Simulation

In this section, we describe a simple computer simulation that supports the plausibility of the model outlined in Section 2.1. In particular, we consider how the fundamental elements of our theory (the innate template, the data input, and the rate of decay and fixation) might interact dynamically with each other to produce autistic conditions.

As emphasized repeatedly in the previous sections, one of the primary differences between a typical and an autistic network—and the one on which we based most of the explanations in Section 3—lies in the nature of the data units that reach fixation. We suggested that under normal circumstances, due to coalescing and segmentation, there are relatively few nodes and these are associated with relatively short data sequences, while under autistic conditions, nodes are typically associated with relatively long data sequences. We further suggested that the primary cause for these segmentation problems in autism is an atypical innate network. We simulate here the basic processes of creating labeled nodes as a result of segmentation, and changing weights of the nodes, as described in Section 2.1, and use this to test whether having an inappropriate initial network can lead to segmentation problems and their consequences, as described in Section 2.2. In our simulation, we ignore the process of link formation and "higher-order" processes like generalization and coalescing, since we want to focus on the number of nodes and the size of the data sequences associated with each node. As we shall see, there are significant qualitative differences between "normal" conditions and "autistic" conditions along these dimensions.

4.1 Data Acquisition

In our simulation, the only data that can be acquired are the following 25 sequences of three letters:

abc	aei	mki	ahc	dbf
def	dhl	jhf	dkf	gei
ghi	gko	gec	gni	jhl
jkl	jnc	dbo	jel	mko
mno	mbf	anl	mho	ani

Table 1: The 25 data sequences used.

Letters can denote words, objects, or part of a visual picture. For example, the sequence "abc" can represent the sentence "Jon-drinks-milk" (or the visual scene of "Jon", "a human hand holding an object", and "a glass of milk"), and the sequence "ahc" might represent a sentence like "Jon-spills-milk".

To keep things simple and tractable, we considered only three-letter data sequences. The simulation could easily be extended to data sequences of variable length; we do not expect the results to be qualitatively different. The size of the simulation data set is obviously small in comparison to any real data set but, again, we do not expect a larger set to lead to qualitative differences.

The simulation uses two parameters:

- γ—this is used as both the initial weight of a new node added to the network and the amount that the weight of a node is increased when the data sequence labeling that node is observed. (Intuitively, a node has weight 0 initially, so the initial weight can be viewed as the increase from a weight in 0 when a data sequence is first observed.)

- δ—this is the amount that the weight of a node decreases at a given step if the node is not reinforced. This parameter corresponds to the decay rate in our earlier discussion.

We run the simulation with different values of γ and δ, to see the effect of these parameters on the outcome.

In Section 2.1 we were deliberately vague as to exactly how data sequences were segmented. For our computer simulation, we have to make some further assumptions. Roughly speaking, we assume that a new data sequence s is segmented if (a) another sequence s' is encountered such that s and s' have a common subsequence s'', and (b) the weight of either s' or s'' is greater than some minimal threshold that we call the *breaking threshold*. We take the fixation threshold to be 1; a node's label is not segmented further once the node's weight reaches 1. Using the breaking threshold ensures that a new sequence must be observed a number of times before it affects the network structure. We considered breaking thresholds of both 0 and 0.5. A threshold of threshold of 0 amounts to running the threshold without a breaking threshold. We believe that, in practice, there is likely to be a positive breaking threshold, both to minimize errors and to avoid the processing of unimportant data. We present the algorithm and report the results for breaking threshold 0.5 here; the results for 0 are qualitatively similar, and are available upon request.

In more detail, we proceed as follows. We assume that at each step in the simulation, the "network" consists of a set of nodes, each of which has an associated weight and is labeled by some data sequence, but no links. (We put "network" in quotes here because there are in fact no links.) Each data sequence is the label of at most one node. Some of the nodes are "active"; the rest are "passive". Passive nodes do not have weights. They can be thought of as dummy nodes, which serve as a technical device for representing data sequences that had weight above the breaking threshold but were broken up into subsequences. There are no passive nodes in the model discussed in Section 2.1; their role is played by links. For example, if "abc" is broken up into "a", "b", and "c", we would expect there to be links with high weight between "a" and "b" and between "b" and "c" (and perhaps also between "a" and "c"). Thus, even if there is no node labeled "abc" in the network, the links provide a way of representing the fact that there was once a node labeled "abc" in the network, and that its subsequences are still in the network. In the simplified setting of our simulation, we can use a passive node labeled "abc" to capture this relationship instead of using a link. Given this intuition, for every passive node in the network, there must be some active nodes labeled with subsequences of the passive node's label. The practical importance of passive nodes in our simulation is that if a sequence such as "abc" has already been seen and segmented before, a newly observed "abc" does not need to wait again until it reaches the breaking threshold before being segmented (see step 4 below, as opposed to step 3).

At each step of the simulation, a single data sequence (e.g. "abc" or "def") is selected at random from those given in Table 1. Thus, we are implicitly assuming that all data sequences are equally likely to be observed. However, some data sequences may be "ignored", in the sense that they do not affect segmentation or the weights of nodes. The likelihood of a data sequence being ignored depends on the nodes already in the network.

In the algorithm, we use the phrases *partial match* and *exact match*. A data sequence x partially matches another data sequence y if y is a subsequence of x; x is an exact match of y if $x = y$. Thus, "abc" partially matches all of "a", "b", "c", "ab", or "bc"; "abc" is an *exact match* only to "abc".

1. If the new data sequence does not match (exactly or partially) the label of any active node in the network, then a new node is added with probability 0.005, labeled with the new data sequence. The new node gets weight γ. Note that we allow new data to be acquired even if it has no partial match in the network, but at a very low probability. Thus, if the data sequence selected is "abc", and "abc" does not (partially) match any node already in the network, then with probability 0.995, it is ignored altogether, and with probability 0.005, a new node is added to the network labeled "abc" and is assigned weight γ. There is evidence that even autistic individuals may acquire some data by chance, for example, if it is heard or observed at a time of another stimulating event [Frith 1989, p. 125]. More generally, mechanisms of temporal association known from associative learning (e.g. [Pearce and Bouton 2001]) may facilitate the sporadic acquisition of data that does not match data already in the network.

2. If the new data sequence exactly matches the label of an active node already in the network, then the weight of that node increases by γ. If, as a result, this weight is greater than or equal to the 0.5 breaking threshold, go to instruction 5; otherwise, go to instruction 6.

3. If the new data sequence does not exactly match the label of either a passive or active node, but it does partially match the label of an active node in the network, add a new node to the network with weight γ, labeled with the new data sequence. (Note that we do not have to check if the new data sequence partially matches a passive node, because if it does, it must also partially match an active node in the network.) Go to instruction 6.

4. If the new data item exactly matches the label of a passive node already in the network (which means that this item has been acquired before), search the network for active nodes with weight greater than or equal to the 0.5 breaking threshold labeled by data sequences that either are subsequences of the new data sequence or overlap with the new sequence. To see what happens then, suppose that the new data sequence is "abc", there is a passive node labeled "abc", and the only active nodes with weight above threshold labeled by partial matches to "abc" have labels "c" and "aei", with weights 1.4 and 0.6, respectively. One of these two nodes is chosen, with a probability according to their relative weights;[12] that is, the node labeled "c" is chosen with probability $1.4/(1.4 + 0.6) = 0.7$; similarly, the node labeled "aei" is chosen with probability 0.3. If the node labeled "c" is chosen, then "abc" is broken up into "ab", and "c". The weight of the nodes labeled "ab" and "c" are both increased by γ. (If there is no node labeled "ab", then one is added to the network with initial weight γ.) Similarly, if "aei" is chosen, then the sequences are broken

[12] In the simulation, for convenience, we use relative weights to choose among alternative segmentation possibilities. We do not mean to suggest that this is actually how people choose among segmentation possibilities in practice. As we mentioned earlier in the paper, we believe that, in a real network, such decisions are often based on context.

up into “a”, “ei”, and “bc”. The new weight of “a” is the sum of its old weight (0 if there was no node labeled “a”, but note that there may be a node labeled “a” in the network with weight below threshold), 0.9 (the old weight of “aei”), and γ (representing the support of the observation “abc” for “a”). The new weight of “ei” is the sum of its old weight and 0.9; the weight of “bc” is the sum of its old weight and γ. Finally, “aei” becomes a passive node. It is important here that “aei” must have weight between 0.5 and 1. If its weight is greater than one, it has become fixated, and then cannot be broken up into subsequences. A sequence such as “ei” that has become fixated can be used to break up larger sequences, but cannot be broken up itself.

5. The network is searched for further pairs of data sequences that can be segmented, using the rules of instruction 4. For example, if the weight of the node labeled “abc” reached the breaking threshold as a result of applying instruction 2, then “abc” can be segmented if a node labeled “aei” is already in the network and has weight above the breaking threshold. Similarly, if a node labeled “ei” reached the breaking threshold as a result of applying instruction 4 and there is a node in the network labeled “gei” with weight above the breaking threshold but not fixated, then it is segmented into “g” and “ei”. This instruction is repeated until no further sequences can be segmented.

6. All nodes with weight < 1 (i.e., below the fixation threshold) that were not observed (i.e., whose weight did not increase as a result of applying instructions 1–5) have their weight decreased by δ. Moreover, if there is a passive node such that all the subsequences of its label are labels of nodes with weight 0, then the passive node disappears from the network. For example, if there is a passive node labeled “abc”, and every node with label “a”, “b”, “c”, “ab”, or “bc” has weight 0, then the node labeled “abc” disappears from the network.[13]

The algorithm summarized by the flowchart in Figure 3.

To compare the “normal” case to the “autistic” case, we run the simulation with different initial networks. For the normal case, we started with two nodes, labeled “a” and “f” respectively, both with initial weight 1. For the autistic case, we started with two nodes, labeled “abc” and “def”, again both with initial weight 1. We can think of these nodes as initial templates, in the sense of the discussion in Section 2.3. Alternatively, they may result from skewed input data distributions early in life. According to this latter possibility, these initial nodes may not represent the innate templates themselves, but rather the effect of the templates on the initial data. For example, we would expect a newborn to have an initial template that results in her focusing on her mother’s face and voice and the words her mother utters. If an autistic child’s initial template is different, the distribution of words that the autistic child focuses on will be different from that of a normal child, resulting in early networks that are already quite different. For example, an autistic child may ignore almost all words other than the one or two that are uttered in the context of her unusual focus of interest.

We can illustrate this point, and at the same time show that the differences between the normal case and the autistic case are really due to differences in the distribution, by considering a modified version of the simulation where the only difference between the normal case and the autistic case is the input

[13] Given our very small rate of decrease and the fact that we choose data sequences at random, it is very unlikely that nodes that at one point have weight above threshold will have their weight go down to 0. Thus, passive nodes essentially never disappear from the network. However, we can imagine that if the distribution of data sequences changes over time, so some data sequences that at one point were relatively common become rare, then this may be a more significant phenomenon.

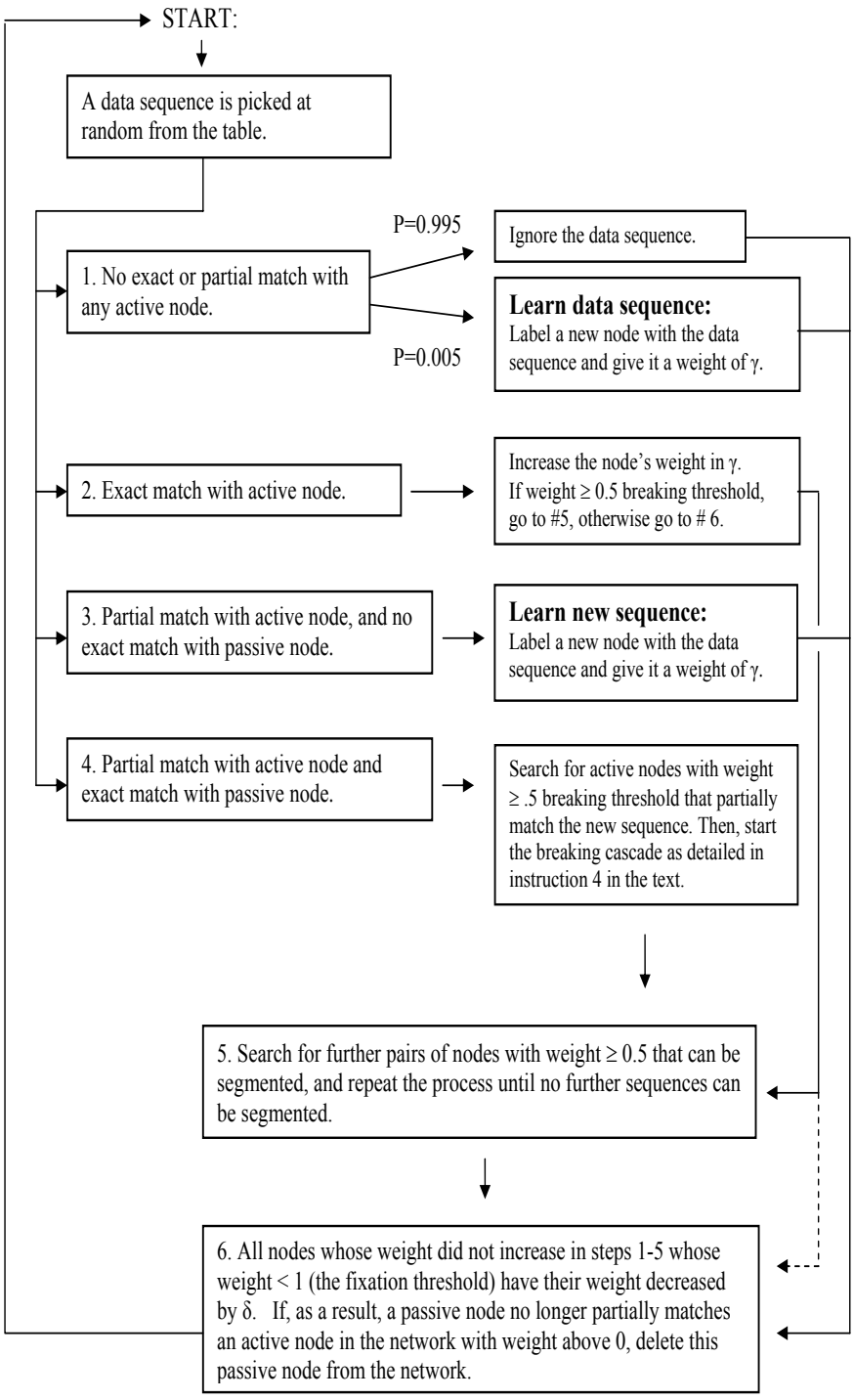

Figure 3: A flowchart describing the algorithm.

distribution. Specifically, we start the same simulation with an empty network (so that there are no initial templates, and all strings have initial probability 0.005 of being acquired in step 1) and consider two input distributions. One is the uniform distribution; the second makes "abc" and "def" 15 times more likely to be chosen than other strings. With the uniform distribution, if we run the simulation sufficiently long (1000 steps suffices), we get normal segmentation. On the other hand, with the skewed distribution, after 500–1000 steps, the first fixed elements are "abc" and "def", and we are essentially back to the case of having "abc" and "def" as the initial template. (This result is robust for all values of γ and δ that we used in our simulation.) This shows that we can simulate the two initial networks that we used by starting with an empty network and taking the initial distribution over data sequences to be either uniform or appropriately skewed.

Of course, we do not need to assume a uniform distribution to get good segmentation. Many other choices will work as well, as long as the initial distribution is not too skewed. For example, an initial distribution that puts an initial weight of 0.495 on "abc" and "def", and splits the remaining 0.01 on the remaining strings, is unlikely to produce good segmentation. Moreover, there is nothing special about this case. It is easy to show analytically that we can simulate the effects of any initial template consisting of nonoverlapping strings by an appropriately skewed distribution.

4.2 Results and Discussion

The results of our simulation for breaking threshold 0.5 are summarized in Figure 4. Recall that there are two parameters in our experiments: the increase γ in the weight of a node when the data matches the label of the node and the amount δ of decay when there is no match. We consider five possible values of γ: 0.1, 0.2, 0.3, 0.4, or 0.5, and five possible values of δ: 0, 0.005, 0.01, 0.015, or 0.02. We took all possible combinations of these two parameters, giving 25 settings. For each setting of the parameters, we considered runs of five possible lengths: 100, 500, 1,000, 5,000, and 10,000 steps. (The length of a run is the number of times we repeat instructions 1–6 of the algorithm above.) For each of the 25 possible (γ, δ) settings and each of the 5 run lengths, we ran the experiment 10 times. We considered the size of fixated sequences at the end of the experiment, averaged over the ten runs where the same setting and the same run length were used, and the number of fixated data sequences, again averaged over the ten runs.

The top graph on the left of Figure 4 considers the average length of fixated units (that is, units with weight ≥ 1) for all 25 (γ, δ) settings when the runs have length 100. The pink squares show the situation for the autistic setting; the blue diamonds show the situation for the normal setting. Each pink square and each blue diamond represents an average of ten runs. Note that, for all the parameter settings, in the autistic case, the fixed units had a length of three. On the other hand, in the normal case, most data sequences are broken up, and the average length of the fixated data sequences is somewhere between 1 and 1.4, depending on the (γ, δ) setting.

The top graph on the right considers how many data sequences become fixated when runs have length 100. In both the autistic case and the normal case, when γ is small and δ is large, the only data sequences that become fixated are the two original data sequences, which started out fixated. On the other hand, as γ gets larger relative to δ, in the normal case, almost all 15 of the data of length 1 get fixated, and perhaps a few of length 2. On the other hand, in the autistic case, no more data sequences get fixated during the first 100 steps.

In the normal case, as runs get longer, almost all data sequences that get fixated continue to have

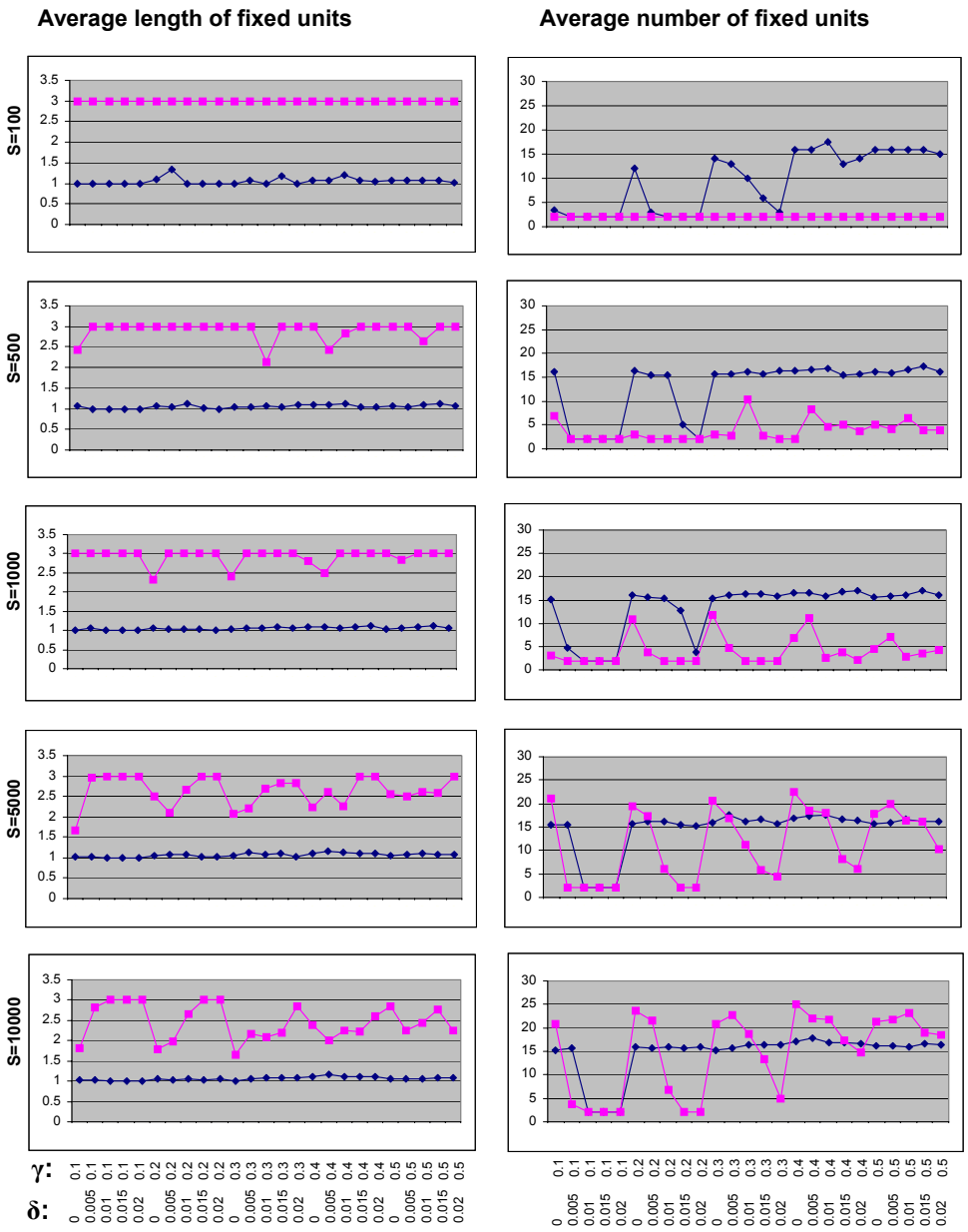

Figure 4: Experimental results.

length 1, and there are more parameter settings for which almost all sequences of length 1 get fixated. The only problematic cases are when $\gamma = 0.1$ (the smallest possible setting) and $\delta \geq 0.01$. We expect that such combinations, where learning is not successful, are unlikely to survive in nature; natural selection will weed them out. To understand why sequences are broken up so efficiently, note that, because the normal case starts with nodes labeled "a" and "f" that are already fixated, if a new data sequence is observed that includes an "a" or an "f" and that data sequence does not already appear in the network, then a new node is added to the network labeled by that data sequence, according to instruction 3 of the algorithm. As long as the ratio of γ to δ is not too small, the weight of that node will increase according to instruction 2 until it reaches threshold. At that point, by instruction 5, it will be broken up. This process very quickly breaks up all data sequences to sequences of length 1. In other words, "a" and "f" serve as "attractors", attracting all new data sequences that include an "a" or an "f" into the network, and breaking them into smaller units, which in turn, attract additional sequences that include these newly broken units (e.g., a broken "ani" previously attracted by "a" can now attract "gni" due to a newly broken "ni"). By way of contrast, "abc" is a bad attractor, since it can only attract other "abc"s.

The autistic case exhibits markedly different behavior. What happens is quite sensitive to the setting of γ and δ. Nevertheless, in almost all cases, the average size of fixated data sequences is above 2. The average number varies considerably. To understand why, consider what is involved in breaking up a subsequence. Note that the "abc" and "def" initially in the network in the autistic case cannot be used to break up anything, since they are already fixated. If an "abc" or "def" is observed, the weight of these nodes increases. On the other hand, if, say, "ahc" is observed, it gets added to the network with probability only 0.005. Thus, in runs of length 100, it is extremely unlikely that any new node gets added, let alone that it is fixated. For a sequence to be broken up into subsequences, two sequences must be added, and the second one to reach threshold must do so before the first gets fixated. This is a rather unlikely event. If runs are sufficiently long, this does happen occasionally, which is why the number of fixed units can be significant in runs of length 1,000 or more, especially if the ratio γ/δ is high. In runs of length 10,000 where the ratio is greater than 30, we often find that many sequences are fixated, but they are long sequences (of which there are many more than the 15 sequences of length 1). Thus, we get a situation where there are more fixated sequences in the autistic case than in the normal case. This is consistent with the possible need for an "enlarged brain" in autistic children, mentioned at the end of Section 3.

To summarize, the computer simulation was able to capture many of the behaviors predicted by our model. As a result of having only a different initial template, in the autistic case, data acquisition was slow, segmentation did not work properly, and the nodes that were fixated were typically labeled by long data sequences. Moreover, learning based on the abnormal initial template was very sensitive to the settings of γ and δ. Roughly speaking, the reason for this is that nodes labeled with longer sequences can attract fewer data sequences ("aei", for example, can attract only "aei", while "ei" can attract "aei" and "gei", and "a" can attract all the five sequences in Table 1 that include an "a"). Attracting more sequences results in more segmentation, which in turn attracts more sequences. If fewer sequences are attracted, the ones in the network are more likely to decay before reaching fixation. As we can see in Figure 4, learning based on the abnormal initial template was indeed successful only when γ was relatively high and δ relatively low.

This observation may provide an additional important insight into the developmental process of autism. Autistic individuals are known to exhibit marked contrasts of performance in different cognitive

tasks (see, e.g., [Baron-Cohen and Bolton 1993; Happé 1999]). One way to explain this phenomenon, as we pointed out earlier, is in terms of inappropriate initial templates, which cause autistic individuals to acquire a great deal of data in some domains but very little or no data in other domains. The simulation results presented here suggest an additional explanation. The great sensitivity to the weight increase/decrease ratio shown by our simulation means that natural fluctuations in this ratio that are unlikely to affect normal children can be critical for autistic children. It is well known that the strength of memorizing or forgetting events or scenes (represented by the increase and decrease rate, respectively, in our simulation) is not uniform, and likely to be affected by the context or the strength of reinforcement [Domjan 1998] and its importance for survival value [Nairne et al. 2007]. In other words, the different increase/decrease ratios considered by our simulation may all be represented in a single individual but for different types of data. Our simulation shows that the "autistic template" response to this variation is to produce peaks and valleys in learning performance, similar to those observed in autistic children.

We believe that we could have generated similar behavior by using different breaking thresholds for the autistic and normal case rather than by using different initial networks. For example, if we took the breaking threshold in the autistic case to be 0.9 instead of 0.5 (but kept it at 0.5 for the normal case), then the small window until fixation in the autistic case would guarantee that very few sequences would be broken up. (We could achieve the same effect by keeping the breaking threshold at 0.5 and lowering the fixation threshold to 0.6.) In any case, we feel that having a smaller window is less consistent with the primary symptoms of autism, such as problems with joint attention and social orientation. Furthermore, a smaller window is likely to lead to a systematic failure of learning in *all* domains. This contradicts the observation that autistic individuals are often able to learn quite well in some domains [Happé 1999; Snyder and Mitchell 1999]. We therefore find it unlikely that autism emerges simply as a result of such a "narrow window".

Finally, we note that our simulation assumed a particular segmentation process. While we have not considered other segmentation mechanisms in detail, we believe that all current approach to segmentation will lead to qualitatively similar results. This is because they do segmentation based on cues that depend on the relative distribution of data units, in particular, the co-occurrence of data units. If a string like "abc" is very common, while none of its components are, all approaches will fail to segment it. (See [Brent 1999] for an overview of current approaches to segmentation.)

5 Conclusions

We have presented a data-driven conceptual model for learning. We have deliberately left many of the details of the model unspecified; they will clearly need to be fleshed out. It may well be the case that some of the specific mechanisms we have suggested are not quite right. Moreover, there are many important representation issues that our naive model does not address, including the representation of goals and hypothesis testing. While we believe that our model could be extended to handle such more sophisticated issues, doing so is beyond the scope of this paper. Despite the shortcomings of the model, we believe that it is a useful tool for explaining autism. In particular, we have shown how an inappropriate distribution of data can result in all the major observations associated with autism. In addition, we have shown how we can use the model itself to explain the inappropriate distribution as the result of an inappropriate initial (or early) network. We believe that the consistency between the symptoms of autism and the data-acquisition disorder predicted by our model suggests that the model itself is correct.

An obvious question at this point is whether our model makes any testable predictions. In fact it does. We consider two predictions here. The most straightforward prediction is that the data-input distribution of a young child can be manipulated in such a way as to lead to autistic behaviors. Of course, this obvious test is unethical (and technically difficult to carry out in any case, since it is hard to force people to pay attention to only certain things).[14] Alternatively, it may be possible to ethically test the effect of more modest "small scale" manipulations. For example, in some studies on word segmentation, sequences of an artificial language were broadcast to human infants (e.g., [Aslin et al. 1998; Gomez 2002; Saffran et al. 1996]). Similar methods may be used to test how different data distributions affect segmentation. Perhaps these experiments could then be extended to study how different data distributions affect a child's representation and ability to predict. (In our model, this amounts to studying how different data distributions affect the construction of the network representation.)

A second prediction made by our model is with regard to the effectiveness of early-intervention methods. While there is currently no medical treatment for autism, it has become increasingly clear that early behavioral intervention is highly beneficial for autistic children [Green 1996; Jensen and Sinclair 2002; Rogers 1998; Volkmar and Pauls 2003]. Indeed, some experts argue that intensive behavioral intervention can even result in a complete recovery from autism [Lovaas 1987; McEachin et al. 1993]. So far, there has been no theory that explains how a behavioral treatment can possibly cure autism, nor why these treatments are successful in some cases but not in others [Bibby et al. 2001; Lovaas 1987]. The claims of possible cures remain controversial, and many experts view behavioral intervention as only one way to facilitate compensatory learning (see, e.g., [Frith 2001], [Howlin 1998, pp. 77-81]). Our model predicts that any treatment that helps autistic children to get a more typical flow of data will reduce the severity of the disorder. Moreover, we can test whether more of a focus on the data distribution can produce significant improvement. Despite differences among early-intervention methods, all of them can be viewed as attempting to finding ways to provide the child with more data, according to a more typical distribution. (Although the treatment methods typically do not speak in terms of data—they speak in terms of "experiences", "communication cycles", "games", "learning", "operant conditioning", and so on—in terms of our model, what they do can be viewed as pushing data.) For example, some behavioral treatments start by using reinforcement to teach a child to create eye contact [Lovaas 1981]. The aim of this approach is to artificially cause autistic children to gain interest in what should have interested them in the first place. If this is successful, the child should start acquiring a more typical data flow, and may recover from the initial problem. (See also Koegel et al. [2001] for related evidence on pivotal areas in intervention for autism.)

On the other hand, our model also suggests why recovery may not be achieved in all cases and, implicitly, why performing a definitive experiment may be difficult. With some severely autistic children it may not be possible to repair the initial network so as to get a more appropriate data flow. And even if the repair process is initially successful, it may not be sufficiently fast and comprehensive to prevent the atypical fixation of long data sequences in the network. In such cases, a child may exhibit substantial

[14]There is anecdotal evidence supporting this prediction, showing that severe social deprivation early in life may result in some features that resemble autism (e.g. [O'Connor and Rutter 2000]). The data distribution for these socially deprived children is unlikely to have the second property that we discussed in Section 2.2 (that certain phrases are encountered frequently but their constituents are not encountered in other contexts), which is why we would not expect them to exhibit exactly the same behavior that we see in autistic children. There is also evidence that socially deprived children make considerable progress after being treated and integrated into society (see Frith's [1989] discussion of the "Kaspar Hauser" case). This too is consistent with our theory; if we assume that their initial network is normal, once they encounter a more normal data distribution, things improve.

gains, but still exhibit some symptoms of autism, even in adulthood. Finally, it is possible that early intervention may not work for a type of autism for which the main problem in data acquisition is not related to the innate network, but rather to a problem in the mechanism that adjusts the weights or the reward value of data sequences.

Nevertheless, by highlighting the critical window between decay and fixation, our model may suggest possible improvements in early-intervention programs. Because issues of timing and the quantity of data are critical for dealing with decay and fixation, the task of restoring an appropriate data flow in early-intervention programs is not simple. Intensive treatment alone may not guarantee the right quantitative balance for obtaining the best results. We suggest that special attention be given to the frequency and the distribution of data sequences provided in early-intervention programs. While the exact solutions need to be studied experimentally with intervention experts, the main principle may be quite simple. Considering the time window between decay and fixation, a child needs to receive data sequences that have sufficient commonalities that they can be broken into natural subsequences. This should help in enabling an autistic child to build a meaningful network.

The biological causes of autism are complex [Eigsti and Shapiro 2003], and although prevention of autism through prenatal genetic screening would be the ideal solution, there is no sign that this will be feasible in the very near future. For the time being, behavioral programs will continue to be the main treatment for autism [Volkmar and Pauls 2003]. We hope that our model can help in improving such programs by providing a theoretical framework in which the effect of quantitative issues like data distribution, decay, and fixation can be tested experimentally. More importantly, if our model is correct, it means that, at least in principle, autism is curable. Even if the model is correct though, it is clear that much further work needs to be done to elucidate the details. We hope that our results have provided some motivation for further research along these lines.

Our model also suggests other lines of research. For one thing, we have stressed the importance of the initial network. The form of the initial network presumably has a biological basis. Perhaps different areas in the brain contain, or are linked to, parts of the original innate network. The innate nodes attract new data sequences and store them in their close vicinity, or in other easily linked areas. Accordingly, a certain area of the brain may become specialized in, for example, face processing, because it is the default storage area for data associated with the nodes that form the template for human faces. There may be other areas in the brain that are responsible for tuning the reinforcement value of data sequences (i.e., the weights of the nodes associated with them), rather than storing them. (It has been suggested that the amygdala serves such a function [Grelotti et al. 2002].) If this model is correct, then a major challenge for future research would be to try to understand better the initial structure of the network and its relation to brain structure. This would then lead to a better understanding of how problems in the initial structure lead to autism.

It would also be of interest to investigate whether brain neuroanatomical abnormalities in autistic individuals are a secondary outcome of developing an atypical network as a result of having an abnormal initial template. For example, we suggested at the end of Section 3 that evidence for the enlarged brain sometimes associated with autism may be explained by the presence of many nodes labeled by long data sequences in an autistic child's network, and that the local overconnectivity and long-range underconnectivity exhibited by the brains of autistic children may also be a consequence of the poor connectivity predicted by our model. Another example comes from a recent study about the lack of the fusiform face area in the cortex of autistic individuals [Grelotti et al. 2002]. Interestingly, Grelotti et al. suggest that the problem is not that a postulated module for face processing is impaired in autistic

individuals, but rather that autistic individuals fail to develop cortical face specialization as a result of their initial reduced interest in human faces.

The possible implications of such innate nodes may go well beyond autism. As we suggested in Section 2.3, variations in cognitive style and talents may be due to slight variations in the structure of the innate initial network. Faced with the same environment, individuals with slightly different initial network may end up having quite different networks in later life. For example, one may be highly interested in visual scenes and will therefore have an extensive network of visual nodes. This, in turn, helps him in storing and using further new data associated with this already rich and useful visual context. Another individual may be initially more interested in sounds and verbal utterances and may develop a relatively richer network of nodes that store verbal information, sounds, or even music. We speculate that much of the variation in what we perceive as intelligence does not reflect variation in basic processing speed, or memory capacity, but rather variations in the structure of the network. Some networks may be better at supporting cognitive skills that we perceive as "intelligent" or "high level", while others may support skills that are no less impressive from a computational point of view, but may be related to more basic motor or visual skills.

We conclude with one final potential line of research: We have viewed autism as the result of one particular set of problems with a child's network. But clearly these are not the only kinds of problems that can occur in network formation; other diseases may be the result of a different set of problems. In particular, we are hoping to investigate the extent to which schizophrenia may be the result of having a network that is "too connected".[15]

Acknowledgments

We thank Simon Baron-Cohen, Matthew Belmonte, Bernard Crespi, Shimon Edelman, Barbara Finlay, Sally Ozonoff, and Michael Spivey for their perceptive comments on the paper and encouragement of the project. Additional comments were received from Amir Ayali, Elhanan Borenstein, Elizabeth Adkins-Regan, Jeff Elman, Sagi Goldman, Zach Solan, and Roy Wollman. Sagi Goldman and Roy Wollman also programmed the simulation and gave many useful comments and suggestions on how it could be improved.

References

Anzalone, M. E. and G. G. Williamson (2000). Sensory processing and motor performance in autism spectrum disorders. In A. M. Wetherby and B. M. Prizant (Eds.), *Autism Spectrum Disorders: A Transitional Developmental Perspective*. Baltimore, Maryland: Paul H Brookes.

Aslin, R. N., J. R. Saffran, and E. L. Newport (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science 9*(4), 321–324.

Aylward, H., N. J. Minshew, K. Field, B. F. Sparks, and N. Singh (2002). Effects of age on brain volume and head circumference in autism. *Neurology 59*, 175–183.

Badcock, C. R. and B. J. Crespi (2006). Imbalanced genomic imprinting in brain developyment: an evolutionary basis for the aetiology of autism. *Journal of Evolutionary Biology 19*, 1007–1032.

[15] We thank Bernard Crespi for this suggestion.

Baron-Cohen, S. (2000). Theory of mind and autism: a fifteen year review. In S. Baron-Cohen, H. Tagar-Flusberg, and D. J. Cohen (Eds.), *Understanding Other Minds, Volume A*, pp. 3–20. Oxford, U.K.: Oxford University Press.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences 6*, 248–254.

Baron-Cohen, S., D. A. Baldwin, and M. Crowson (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development 68*, 48–57.

Baron-Cohen, S. and P. Bolton (1993). *Autism: The Facts*. Oxford, U.K.: Oxford University Press.

Baron-Cohen, S., A. Leslie, and U. Frith (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology 4*, 113–125.

Baron-Cohen, S., J. Richler., D. Bisarya, N. Gurunathan, and S. Wheelwright (2003). The systemizing quotient: an inxvestigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Transactions of the Royal Society: Biological Sciences 358*(1430), 361–374.

Bateson, P. (1966). The characteristics and context of imprinting. *Biological Reviews 41*, 177–220.

Bateson, P. (1979). How do sensitive periods arise and what are they for? *Animal Behavior 27*, 470–486.

Belmonte, M. K., G. Allen, A. Beckel-Mitchener, L. M. Boulanger, R. A. Carper, and S. J. Webb (2004). Autism and abnormal development of brain connectivity. *J. Neuroscience 24*(2), 9228–9231.

Bibby, P., S. Eikeseth, N. T. Martin, O. C. Mudford, and D. Reeves (2001). Progress and outcomes for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities 22*, 425–447.

Birch, S. A. J. and P. Bloom (2001). Children are cursed: an asymmetric bias in mental-state attribution. *Psychological Science 14*, 283–286.

Bloom, P. (2001). Précis of how children learn the meanings of words. *Behavioral and Brain Sciences 24*, 1095–1103.

Brachman, R. (1985). "I lied about the trees" (or, defaults and definitions in knowledge representation). *AI Magazine 6*(3), 80–93.

Brent, M. R. (1999). Speech segmentation and word discovery: a computational perspective. *Trends in Cognitive Sciences 3*(8), 294–301.

Brian, J. A. and S. E. Bryson (1996). Disembedding performance and recognition memory in autism. *J. Child Psychology and Psychiatry 37*, 865–872.

Brock, J., C. Brown, J. Boucher, and G. Rippon (2002). The temporal binding deficit hypothesis of autism. *Development and Psychopathology 14*(2), 209–224.

Bushwick, N. L. (2001). Social learning and the etiology of autism. *New Ideas in Psychology 19*, 49–75.

Carpenter, M. and M. Tomasello (2000). Joint attention, cultural learning, and language acquisition: Implications for children with autism. In A. M. Wetherby and B. M. Prizant (Eds.), *Autism*

Spectrum Disorders: A Transitional Developmental Perspective. Baltimore, Maryland: Paul H Brookes.

Cave, J. M. W. K. R. and S. L. Franzel (1989). A modified feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance 15*, 419–433.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Cohen, I. L. (1994). An artificial neural network analogue of learning in autism. *Biological Psychiatry 36*, 5–20.

Courchesne, E., C. M. Karns, H. Davis, R. Ziccardi, R. Carper, Z. Tigue, H. Chisum, P. Moses, K. Pierce, C. Lord, A. Lincoln, S. Pizzo, L. Schreibman, R. Haas, N. Akshoomoff, and R. Courchesne (2001). Unusual brain growth patterns in early life in patients with autistic disorder: An MRI study. *Neurology 57*, 245–254.

Courchesne, E. and K. Pierce (2005). Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current Opinion in Neurobiology 15*, 225–230.

Courchesne, E., E. Redcay, and D. P. Kennedy (2004). The autistic brain: birth through adulthood. *Current Opinion in Neurology 17*(4), 489–496.

DeCasper, A. and W. P. Fifer (1980). Of human bonding: Newborns prefer their mother's voice. *Science 208*, 1174–1176.

Domjan, M. (1998). *The Principles of Learning and Behavior*. Pacific Grove, Calif.: Brooks/Cole.

Eigsti, I. M. and T. Shapiro (2003). A systems neuroscience approach to autism: biological, cognitive, and clinical perspectives. *Mental Retardation and Developmental Disabilities Research Reviews 9*, 205–215.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science 14*, 179–211.

Elman, J. L. (1999). Origins of language: a conspiracy theory. In B. MacWhinney (Ed.), *The Emergence of Language*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Emerich, D. M., N. A. Creaghead, S. M. Grether, D. Murray, and C. Grasha (2003). The comprehension of humorous materials by adolescents with high-functioning autism and Asperger's syndrome. *Journal of Autism and Developmental Disorders 33*(3), 253–257.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of gaze. *Neuroscience and Biobehavioral Reviews 24*, 581–604.

Emery, N. J. and N. S. Clayton (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature 414*, 443–446.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995). *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.

Fahlman, S. (1979). *NETL: A System for Representing Real-World Knowledge*. Cambridge, Mass.: MIT Press.

Frith, U. (1989). *Autism: Explaining the Enigma*. Oxford, U.K.: Blackwell.

Frith, U. (2001). Mind blindness and brain in autism. *Neuron 32*, 969–979.

Frith, U. and M. Snowling (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology 1*, 329–342.

Gallagher, H. L. and C. D. Frith (2003). Functional imaging of theory of mind. *Trends in Cognitive Sciences 7*, 77–83.

Garner, J. P., C. L. Meehan, and J. A. Mench (2003). Stereotypies in caged parrots, schizophrenia and autism: Evidence for a common mechanism. *Behavioural Brain Research 145*, 125–134.

Gobet, F., P. C. R. Lane, S. Croker, P.-H. Cheng, G. Jones, I. Oliver, and J. M. Pine (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences 5*(6), 236–243.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science 13*, 431–436.

Green, G. (1996). Early behavioral interventions for autism: What does the research tell us? In C. Maurice (Ed.), *Behavioral Intervention for Young Children with Autism*. Austin, TX: Pro-Ed.

Grelotti, D. J., I. Gauthier, and R. T. Schultz (2002). Social interest and the development of cortical face specialization: What autism teaches us about face processing. *Developmental Psychobiology 40*, 213–225.

Gustafsson, L. (1997). Inadequate cortical feature maps: A neural circuit theory of autism. *Biological Psychiatry 42*, 1138–1147.

Gustafsson, L. and A. P. Paplinski (2004). Self-organization of an artificial neural network subjected to attention shift impairments and familiarity preference, characteristics studied in autism. *J. Autism and Developmental Disorders 34*, 189–198.

Haith, M. M., T. Bergman, and M. J. Moore (1977). Eye contact and face scanning in early infancy. *Science 198*, 853–855.

Happé, F. (1999). Autism: Cognitive deficit or cognitive style. *Trends in Cognitive Sciences 3*(6), 216–222.

Happé, F. and U. Frith (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism and Developmental Disorders 36*, 5–25.

Hasson, O. (1991). Pursuit-deterrent signals—communication between prey and predator. *Trends in Ecology and Evolution 9*, 325–326.

Hauber, M. E., S. A. Russo, and P. W. Sherman (2001). A password for species recognition in a brood-parasitic bird. *Proc. Royal Society of London Series B 268*, 1041–1048.

Hermelin, B. (2001). *Bright Splinters of the Mind: A Personal Story of Research With Autistic Savants*. London: Jessica Kingsley.

Hermelin, B. and N. O'Connor (1967). Remembering of words by psychotic and subnormal children. *British Journal of Psychology 58*(3/4), 213–218.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences 5*, 253–261.

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences 8*, 26–32.

Hobson, R. P., J. Ouston, and A. Lee (1988). What's in a face? The case of autism. *British Journal of Psychology 79*, 441–453.

Howlin, P. (1998). *Children with Autism and Asperger Syndrome: A Guide for Practitioners and Carers*. New York: John Wiley and Sons.

Insel, T. R., D. J. O'Brien, and J. F. Leckman (1999). Oxytocin, vasopression, and autism: Is there a connection? *Biological Psychiatry 45*, 145–157.

Jensen, V. K. and L. V. Sinclair (2002). Treatment of autism in young children: Behavioral intervention and applied behavior analysis. *Infants and Young Children 14*(4), 42–52.

Johnson, M. H. (2000). Cortical specialization for higher cognitive functions beyond the maturational model. *Brain and Cognition 42*, 124–127.

Jolliffe, T. and S. Baron-Cohen (2001a). A test of central coherence theory: Can adults with high-functioning autism or Asperger syndrome integrate fragments of an object? *Cognitive Neuropsychiatry 6*(3), 193–216.

Jolliffe, T. and S. Baron-Cohen (2001b). A test of central coherence theory: Can adults with high-functioning autism or Asperger syndrome integrate objects in context? *Visual Cognition 8*(1), 67–101.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences 3*(9), 323–328.

Kacelnik, A. and M. Bateson (1997). Risk-sensitivity: Cross-roads for theories of decision-making. *Trends in Cognitive Sciences 1*, 304–309.

Kahn, R. M. and M. A. Arbib (1973). A cybernetic approach to childhood psychosis. *Journal of Autism and Childhood Schizophrenia 3*(3), 261–273.

Kleinberg, J. (2001). Small-world phenomena and the dynamics of information. In T. G. Dietterich, S. Becker, and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, pp. 431–438. MIT Press.

Klin, A. (91). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *J. Autism and Developmental Disorders 21*(1), 29–42.

Klin, A., W. Jones, R. Schultz, and F. Volkmar (2003). The enactive mind, or from action to cognition lessons from autism. *Philosophical Transactions of the Royal Society: Biological Sciences 358*(1430), 345–360.

Koegel, R. L., L. K. Koegel, and E. K. McNerney (2001). Pivotal areas in intervention for autism. *Journal of Clinical Child Psychology 30*(1), 19–32.

Krebs, J. R. and R. Dawkins (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs and N. B. Davies (Eds.), *Behavioral Ecology: An Evolutionary Approach*, pp. 380–402. Oxford, U.K.: Blackwell Scientific.

Leslie, A. (1987). Pretence and representation: The origins of "theory of mind". *Psychological Review 94*, 412–426.

Lin, L., R. Osan, and J. Z. Tsien (2006). Organizing principles of real-time memory encoding: neural clique assemblies and universal neural codes. *Trends in Neurosciences 29*(1), 48–57.

Lopez, B. and S. R. Leekam (2003). Do children with autism fail to process information in context? *J. Child Psychology and Psychiatry 44*, 285–300.

Lotem, A. (1993). Learning to recognize nestlings is maladaptive for cuckoo cuculus canorus host. *Nature 362*, 743–745.

Lovaas, O. I. (1981). *Teaching Developmentally Disabled Children: The ME Book*. Baltimore, Maryland: University Park Press.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Clinical and Consulting Psychology 55*, 3–9.

McClelland, J. L. (2000). The basis of hyperspecificity in autism: A preliminary suggestion based on properties of neural nets. *J. Autism and Developmental Disorders 30*, 497–502.

McClelland, J. L. and T. T. Rogers (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience 4*, 1–14.

McEachin, J., T. Smith, and O. Lovaas (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation 97*, 359–372.

Miller, R. A., N. Kleinhans, N. Kemmotsu, K. Pierce, and E. Courchesne (2003). Abnormal variability and distribution of functional maps in autism: An fMRI study of visuomotor learning. *American Journal of Psychiatry 160*, 1847–1862.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Review 81*, 521–535.

Mottron, L., J. A. Burack, J. E. A. Stauder, and P. Robaey (1999). Perceptual processing among high-functioning persons with autism. *J. Child Psychology and Psychiatry 40*, 203–211.

Mottron, L., M. Dawson, I. Soulieres, B. Hubert, and J. A. Burack (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *J. Autism and Developmental Disorders 36*, 27–43.

Mundy, P. (1995). Joint attention and social-emotional approach behavior in children autism. *Development and Psychopathology 7*, 63–82.

Nairne, J. S., S. R. Thompson, and J. N. S. Pandeirada (2007). Adaptive memory: survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition 33*(2), 263–273.

O'Connor, T. G. and M. Rutter (2000). Attachment disorder behavior following early severe deprivation: extension and longitudinal follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry 39*(6), 703–713.

Oliver, A., M. H. Johnson, A. Karmiloff-Smith, and B. Pennington (2000). Deviation in the emergence of representations: A neuroconstructivist framework for analyzing developmental disorders. *Developmental Science 3*(1), 1–40.

O'Riordan, M. and K. Plaisted (2001). Enhanced discrimination in autism. *The Quarterly Journal of Experimental Psychology 54A*(4), 961–979.

Ozonoff, S. (1997). Components of executive function in autism and other disorders. In J. Russell (Ed.), *Autism as an Executive Disorder*, pp. 179–211. Oxford, U.K.: Oxford University Press.

Ozonoff, S., D. L. Strayer, W. M. McMahon, and F. Filloux (1994). Executive function abilities in autism: An information processing approach. *J. Child Psychology and Psychiatry 35*, 1015–1031.

Pearce, J. M. and M. E. Bouton (2001). Theories of associative learning in animals. *Annual Review of Psychology 52*, 111–139.

Pennington, B. F. and S. Ozonoff (1996). Executive function and developmental psycopathology. *J. Child Psychology and Psychiatry 37*(1), 51–87.

Pierce, K., R. A. Muller, J. Ambrose, G. Allen, and E. Courchesne (2001). Face processing occurs outside the fusiform "face area" in autism: evidence from functional MRI. *Brain 124*, 2059–2073.

Pinker, S. (1994). *The Language Instinct*. New York: W. Morrow and Company.

Plaisted, K., V. Dobler, S. Bell, and G. Davis (2006). The microgenesis of global perception in autism. *J. Autism and Developmental Disorders 36*, 107–116.

Pullum, G. K. (1996). Learnability, hyperlearning, and the poverty of the stimulus. In J. Johnson, M. L. Juge, and J. L. Moxley (Eds.), *Proc. 22nd Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Role of Learnability in Grammatical Theory*, pp. 498–513. Berkeley, Calif.: Berkeley Linguistics Society.

Ramnani, N. and C. Miall (2004). A system in the human brain for predicting the actions of others. *Nature Neuroscience 7*(1), 85–90.

Rodriguez-Girones, M. A. and A. Lotem (1999). How to detect a cuckoo egg: A signal detection theory model for recognition and learning. *American Naturalist 153*, 633–648.

Rogers, S. J. (1998). Empirically supported comprehensive treatments for young children with autism. *Journal of Clinical Child Psychology 27*, 168–179.

Ropar, D. and P. Mitchell (1999). Are individuals with autism and Asperger's syndrome susceptible to visual illusions? *J. Child Psychology and Psychiatry 40*, 1287–1293.

Russell, J. (1997). *Autism as an Executive Disorder*. Oxford, U.K.: Oxford University Press.

Saffran, J. R., R. N. Aslin, and E. L. Newport (1996). Statistical learning by 8-month-old infants. *Science 274*, 1926–?

Shah, A. and U. Frith (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry 24*, 613–620.

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge, U.K.: Cambridge University Press.

Shettleworth, S. (1998). *Cognition, Evolution, and Behavior*. New York: Oxford University Press.

Singer, T., B. Seymour, J. O'Doherty, H. Kaube, R. J. Dolan, and C. Frith (2004). Empathy for pain involves the affective but not sensory components of pain. *Science 303*, 1157–1162.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences 7*(6), 246–251.

Snowling, M. and U. Frith (1986). Comprehension in "hyperlexic" readers. *Journal of Experimental Child Psychology 42*, 392–415.

Snyder, A., T. Bossomaier, and J. D. Mitchell (2004). Concept formation: 'Object' attributes dynamically inhibited from conscious awareness. *Journal of Integrative Neuroscience 3*, 31–46.

Snyder, A. W. and D. J. Mitchell (1999). Is integer arithmetic fundamental to mental processing? The mind's secret arithmetic. *Proc. Royal Society of London Series Series B 266*, 587–592.

Soha, J. A. and P. Marler (2000). A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Animal Behaviour 60*, 297–306.

Solan, Z., D. Horn, E. Ruppin, and S. Edelman (2005). Unsupervised learning of natural languages. *Proc. National Academy of Sciences 102*(33), 11629–11634.

Solan, Z., E. Ruppin, D. Horn, and S. Edelman (2003). Automatic acquisition and efficient representation of syntactic structures. In S. Becker, S. Thrun, and K. Obermayer (Eds.), *Advances in Neural Information Processing Systems 15 (NIPS 2002)*, pp. 91–98. Cambridge, MA: MIT Press.

Tager-Flusberg, H., R. Joseph, and S. Folstein (2001). Current directions in research on autism. *Mental Retardation and Developmental Disabilities Research Reviews 7*, 21–29.

Temple, C. (1997). *Developmental Cognitive Neuropsychology*. Hove, East Sussex, U.K.: Psychology Press.

ten Cate, C. and D. R. Vos (1999). Sexual imprinting and evolutionary processes in birds: A reassessment. *Advances in the Study of Behavior 28*, 1–31.

Thomas, M. and A. Karmiloff-Smith (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioral and Brain Sciences 25*, 727–788.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences 4*(4), 156–163.

Treisman, A. and G. Gelade (1980). A feature integration theory of attention. *Cognitive Psychology 12*, 97–136.

Volkmar, F. and D. Pauls (2003). Autism. *The Lancet 362*, 1133–1141.

Wallace, G. L. and D. Treffert (2004). Head size and autism. *The Lancet 363*, 1003–1004.

Zahavi, A. and A. Zahavi (1997). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford, U.K.: Oxford University Press.