A connectionist model of category learning by individuals with high-functioning autism spectrum disorder

Alexander Dovgopoly · Eduardo Mercado III

Abstract Individuals with autism spectrum disorder (ASD) show atypical patterns of learning and generalization. We explored the possible impacts of autism-related neural abnormalities on perceptual category learning using a neural network model of visual cortical processing. When applied to experiments in which children or adults were trained to classify complex two-dimensional images, the model can account for atypical patterns of perceptual generalization. This is only possible, however, when individual differences in learning are taken into account. In particular, analyses performed with a self-organizing map suggested that individuals with high-functioning ASD show two distinct generalization patterns: one that is comparable to typical patterns, and a second in which there is almost no generalization. The model leads to novel predictions about how individuals will generalize when trained with simplified input sets and can explain why some researchers have failed to detect learning or generalization deficits in prior studies of category learning by individuals with autism. On the basis of these simulations, we propose that deficits in basic neural plasticity mechanisms may be sufficient to account for the atypical patterns of perceptual category learning and generalization associated with autism, but they do not account for why only a subset of individuals with autism would show such deficits. If variations in performance across subgroups reflect heterogeneous neural abnormalities, then future behavioral and neuroimaging studies of individuals with ASD will need to account for such disparities.

Keywords Autism · Perceptual categorization · Perceptual learning · Computational model · Neural network · Connectionism · Parallel distributed processing

A. Dovgopoly · E. Mercado III (✉)
Department of Psychology, University at Buffalo, SUNY,
Buffalo, NY 14260, USA
e-mail: emiii@buffalo.edu

Autism spectrum disorder (ASD) is a developmental disorder characterized by a variety of cognitive, social, and behavioral impairments. Among other symptoms, individuals with ASD may exhibit deficits in social interaction, difficulties in verbal and nonverbal communication, repetitive or stereotyped behavior, learning difficulties, and extreme fluctuations in mood (American Psychiatric Association, 2000). Individuals with ASD also show atypical perceptual processing (see Dakin & Frith, 2005, for review), including superior perceptual discrimination in some tasks (Mottron, Dawson, Soulières, Hubert, & Burack, 2006; Plaisted, O'Riordan, & Baron-Cohen, 1998) but difficulties in perceptual learning and generalization in others (Church et al., 2010; Gastgeb, Rump, Best, Minshew, & Strauss, 2009; Klinger & Dawson, 2001; Klinger, Klinger, & Pohlig, 2007). Perceptual deficits, in particular, are thought to reflect abnormalities in cortical structure and function in individuals with ASD (Markram & Markram, 2010; Rubenstein & Merzenich, 2003; Spencer et al., 2000).

Researchers have proposed several explanations for the variations in perceptual processing associated with ASD. Qualitative theories include the ideas of weak central coherence (Frith, 1989; Happé & Frith, 2006) and enhanced perceptual functioning (Mottron & Burack, 2006; Mottron et al., 2006), as well as the reduced perceptual similarity hypothesis (Plaisted, 2001). Neurally based accounts of perceptual deficits have pointed to effects of cortical underconnectivity and disrupted cortical connectivity (Just, Cherkassky, Keller, & Minshew, 2004; Kana, Libero, & Moore, 2011), poor functioning of the dorsal/magnocellular system (Spencer et al., 2000), minicolumn pathology (Casanova, Buxhoeveden, Switala, & Roy, 2002), and an imbalance of neural excitation and inhibition (Rubenstein & Merzenich, 2003). Such qualitative explanations of dysfunction in perception provide plausible accounts of why perceptual processing and learning might differ in individuals with ASD.

Dysfunctional learning and generalization processes can also potentially contribute to atypical perceptual processing in individuals with ASD. Although researchers historically have debated about how autism affects learning processes (reviewed by Dawson, Mottron, & Gernsbacher, 2008), recent studies have shown clear early learning differences relative to typically developing (TD) individuals (Solomon, Smith, Frank, Ly, & Carter, 2011; Yechiam, Arshavsky, Shamay-Tsoory, Yaniv, & Aharon, 2010). Deficits in the generalization of learned perceptual and cognitive skills have long been noted as a feature of ASD (Lovaas, Koegel, & Schreibman, 1979; Schriebman, Koegel, & Craig, 1977; Solomon et al., 2011). The cumulative effects of abnormal learning experiences during development could lead to systematic differences in how individuals with ASD represent perceptual events.

Given that perception and learning processes differ in individuals with autism, one might expect that their perceptual category learning might also be atypical. Consistent with this prediction, several studies have demonstrated atypical prototype formation or categorization by individuals with ASD (Church et al., 2010; Gastgeb, Dundas, Minshew, & Strauss, 2012; Gastgeb et al., 2009; Gastgeb, Wilkinson, Minshew, & Strauss, 2011; Klinger & Dawson, 2001; Klinger et al., 2007). Other studies, however, have reported that perceptual category learning by individuals with ASD is relatively unimpaired (Bott, Brock, Brockdorff, Boucher, & Lamberts, 2006; Froehlich et al., 2012; Molesworth, Bowler, & Hampton, 2005; Soulières, Mottron, Giguère, & Larochelle, 2011; Soulières, Mottron, Saumier, & Larochelle, 2007; Vladusich, Olu-Lafe, Kim, Tager-Flusberg, & Grossberg, 2010). These seemingly contradictory findings may reflect differences in methodology, the ages of the participants, their levels of cognitive ability, or the high variability of perceptual performance across individuals with ASD (Mottron et al., 2006). Alternatively, these findings may reflect interactions between neural deficits and task difficulty or complexity (Samson, Mottron, Jernel, Belin, & Ciocca, 2006).

Past work on category learning by individuals with ASD has focused on possible differences in their abilities to categorize items on the basis of similarities between category members. Researchers initially reported that children with ASD were only able to classify items using rules, showing limited capacity to recognize novel, prototypical stimuli as being members of a learned category (Klinger & Dawson, 2001). Later studies, however, showed that adults with high-functioning ASD (HFASD) were able to recognize prototypes (Bott et al., 2006; Froehlich et al., 2012; Molesworth, Bowler, & Hampton, 2008; Soulières et al., 2011), suggesting that all forms of perceptual category learning were intact in adults. Recently, several laboratories have independently converged on a task for assessing visual category learning by individuals with ASD that involves training individuals

to classify abstract patterns of dots into categories (Church et al., 2010; Froehlich et al., 2012; Gastgeb et al., 2012; Schipul, 2012; Vladusich et al., 2010). This image classification task is a classic method in category-learning research (Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968) that has several advantages, including the following: (a) it uses abstract, unfamiliar shapes or dot patterns, thus controlling for impacts of past experience; (b) the shapes have no direct social relevance; and (c) task performance is not easily improved by applying simple rules. The task has been extensively analyzed (Ashby & Ell, 2001; Nosofsky & Zaki, 1998; Smith & Minda, 2001) and has been used in several neuroimaging studies (Little & Thulborn, 2006; Nosofsky, Little, & James, 2012; Reber, Stark, & Squire, 1998). As in the earlier categorization experiments, some of these studies reported atypical prototype identification by individuals with HFASD (Church et al., 2010; Gastgeb et al., 2012), whereas others reported typical prototype formation (Froehlich et al., 2012; Schipul, 2012; Vladusich et al., 2010).

In the following sections, we first discuss how known neural abnormalities associated with autism may contribute to observed deficits in category learning and how computational models can be used to simulate the effects of such abnormalities. We then describe a series of simulations that clarify which neurally based manipulations satisfactorily explain the deficits in category learning observed in past behavioral experiments, and that shed new light on possible sources of the discrepant findings from past studies of categorization by individuals with autism.

Theoretical approach

Numerous quantitative models of category learning are available that can account for the typical patterns of generalization (reviewed by Pothos & Wills, 2011; Wills & Pothos, 2012), including models that incorporate hypotheses about the neural substrates of perceptual category learning (Ashby & Maddox, 2005, 2010). In principle, any of these quantitative/computational models would be able to generalize appropriately after learning to classify dot patterns, and such models might be able to account for the different generalization patterns observed in individuals with ASD. In practice, however, none of these models appear to have been used to predict how individuals with ASD should generalize in specific perceptual category-learning tasks.

General approach to simulating neural substrates of category learning

Existing neurocomputational models of category learning have traditionally focused on assigning different computational

functions to specific brain regions. Several different brain regions are engaged when typical individuals learn to classify dot patterns, including multiple cortical regions, the hippocampal region, and the basal ganglia (Daniel et al., 2011; Nosofsky et al., 2012). Abnormalities in all of these regions have been reported in individuals with autism (reviewed by Penn, 2006; Schroeder, Desrocher, Bebko, & Cappadocia, 2010), and thus could potentially contribute to atypical category learning. The present simulations focus on visual cortical pathways because (1) processing in these regions is known to be abnormal in individuals with autism (Baruth, Casanova, Sears, & Sokhadze, 2010; Behrmann, Thomas, & Humphreys, 2006; Simmons et al., 2009; Vandenbroucke, Scholte, van Engeland, Lamme, & Kemner, 2008); (2) abnormal processing of visual inputs would necessarily impact all subsequent stages of neural processing; and (3) researchers have theorized that learning to classify dot patterns in tasks such as the one used by Church et al. (2010) is especially dependent on visual cortical regions (Ashby & Maddox, 2005). A basic assumption underlying the present simulations is that learning of new perceptual categories requires adjustments within cortical circuits, and that abnormalities in the neural mechanisms that enable such adjustments can lead to atypical category formation and generalization. This assumption naturally leads to the question of when and how such neural abnormalities will manifest themselves in perception and behavior.

The aim of the present simulations was not to evaluate how well existing neurocomputational models of either category learning or autism can explain perceptual generalization by individuals with ASD. Rather, a "theory-neutral" model of visual cortical processing was tested (described below). This model was not originally designed to account for either autistic deficits or perceptual category-learning phenomena. Using it as the foundation of our simulations made it possible for us to objectively evaluate whether abnormalities in learning and stimulus processing might account for the atypical generalization observed in individuals with autism. A potential concern in any modeling effort is that it is relatively easy to construct and customize a computational model to replicate almost any pattern of data, given a sufficient number of adjustable parameters. The present simulations obviate this problem by using a computational model designed by another laboratory, for a different purpose, as the present model's starting point, and by limiting the adjustment of parameters to increases or decreases of a single parameter per simulation (i.e., no attempt was made to identify an optimal set of parameter settings). This approach greatly constrains any opportunity for customization of the model to fit the data (Wills & Pothos, 2012). Modeling empirical data using basic neural network parameter manipulations provides a means of

quantitatively examining the potential etiological factors involved in ASD.

Simulating dysfunctional neural plasticity and homeostasis
Subcellular structural and functional abnormalities are thought to contribute to the neurological problems associated with autism. For example, genetic studies have indicated that ASD is associated with a greater than typical occurrence of mutation in the genes affecting synaptogenesis and synaptic function (Auerbach, Osterweil, & Bear, 2011; Bourgeron, 2009). Ramocki and Zoghbi (2008) suggested that abnormalities at the genetic level result in dysfunctions in subcellular biological processes that impair synaptic homeostasis, resulting in autistic symptoms. For instance, mutations in synaptic proteins identified in patients with ASD may affect the size and shape of dendritic spines (Bourgeron, 2009) and decrease synaptic transmission. Such mutations can negatively impact the development and stabilization of synapses; in animal models of autism, they are associated with hyper-responsiveness, increased stereotypy, and social deficits (Schmeisser et al., 2012). Synaptic abnormalities also have been linked to the pathogenesis of Rett syndrome, an ASD (Moretti et al., 2006). Finally, abnormal sleep patterns in ASD-related disorders suggest that abnormal synaptic scaling and plasticity may also contribute to autism (Wang, Grone, Colas, Appelbaum, & Mourrain, 2011). Such abnormalities can potentially negatively affect incremental-learning mechanisms at a synaptic level, which might in turn disrupt visual category learning. We simulated the effects of synaptic deficits on visual processing by adjusting a model parameter (learning rate) that determined how rapidly connections between processing units were adjusted and, in another variant of the model, by manipulating a second parameter (weight decay) that specified in what ways connections were adjusted during learning. A basic assumption of this approach is that irregularities in synaptic function will degrade neural plasticity and that decreases in neural plasticity will tend to impair learning.

Simulating cortical structural abnormalities The brains of individuals with autism have a greater than typical number of cortical minicolumns—structures that constitute basic functional assemblies of neurons in the brain (Casanova et al., 2002; Casanova et al., 2006). The availability of cortical minicolumns may influence perceptual category learning by modulating a person's capacity to acquire novel stimulus representations and by constraining the possible resolution of those representations (Mercado, 2008). The greater number of minicolumns in individuals with autism has the potential to improve or degrade performance in perceptual category learning, depending on how these cortical circuits contribute to the processing of visual inputs. To explore whether variations in the number of processing units might

contribute to atypical category learning and generalization by individuals with HFASD, we increased and decreased the number of computational processing units (known as "hidden-layer" units) in the model, and examined how these changes impacted on the model's learning and generalization (see also Cohen, 1994, 1998). In this approach, hidden-layer units are considered to be analogous to cortical mini-columns (or sets of minicolumns), and increases in the number of minicolumns are simulated by increasing the number of hidden-layer units.

Simulating noisy stimulus processing The relative increase in cortical excitation observed in individuals with autism may lead to a decrease of signal relative to noise in neural processes (Rubenstein & Merzenich, 2003; Yizhar et al., 2011). Increased neural noise may effectively reduce the efficacy of visual cortical circuits that are engaged during category learning and may also degrade mechanisms for adjusting the selectivity of cortical processing. In past studies, modelers have simulated a decrease in the neural signal-to-noise ratio by modulating the computations performed within a connectionist model (Fellous & Linster, 1998; Li, von Oertzen, & Lindenberger, 2006). To simulate the postulated signal-to-noise decrease in autism, we similarly adjusted a processing property of individual computational units, known as the "gain" parameter. Decreases in neural signal to noise are thus simulated as decreases in the network's capacity to resolve inputs.

In the present article, we focus on simulating the behavioral data reported by Church et al. (2010), and secondarily on experiments conducted by Vladusich et al. (2010). We consider the findings of these studies to be representative of the kinds of deficits reported in past studies of category learning by individuals with ASD.

Simulation 1

Simulating visual category learning by typically developing children

As the starting point of our simulations, we chose an artificial neural network (ANN) originally developed by Henderson and McClelland (2011) to model visual object perception by individuals with simultagnosia. As in traditional theories of visual perception (Goodale & Milner, 1992), this model includes a pathway associated with object identification (corresponding to the ventral cortical pathway) and another devoted to processing location-relevant information (the dorsal cortical pathway). Additionally, an emergent property of their ANN is that the dorsal pathway facilitates object recognition (as was previously suggested by Konen & Kastner, 2008). In this model, the two

pathways cooperate to facilitate recognition of multiple objects across a spatial plane. We selected this model because Henderson and McClelland reported that reducing the number of processing units in the model led to better spatial generalization, but also to less accurate identification and categorization of multiple objects; Cohen (1994, 1998) noted a similar phenomenon in his computational models of autism.

Description of the model

Architecture and implementation Henderson and McClelland (2011) tested three architectures: a recurrent backpropagation neural network with a dorsal and a ventral pathway connected by a hidden layer; a recurrent ventral-pathway-only network; and a feed-forward dorsal-only network. The architecture of the full ventral–dorsal ANN is illustrated in Fig. 1. The neural network simulations reported in this study were conducted with the PDPTool program (www.stanford.edu/group/pdplab/resources.html#pdptool) running in MATLAB R2010a. Ancillary data processing and formatting were handled with custom scripts written in the Perl 5.12.2, Ruby 1.9.2, and Python 2.7 programming languages.

Pilot studies revealed that for the visual category-learning task used by Church et al. (2010), the dorsal pathway learned much more quickly than the ventral pathway, and there was little interaction between the pathways. Consequently, we chose to focus on the dorsal-only architecture for all simulations but Simulation 4. The performance of TD children was simulated using an ANN with 144 input nodes, 144 hidden-layer nodes, and 144 output-layer nodes, as in the original Henderson and McClelland (2011) dorsal-only simulations; their original parameter settings were also preserved, with the exception of learning rate (i.e., no attempt was made
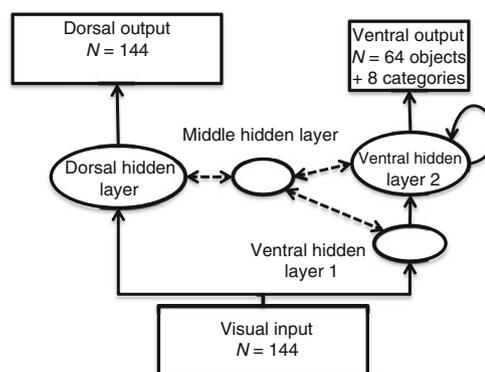
Fig. 1 Henderson and McClelland's (2011) dorsal–ventral object perception model. This architecture was the starting point for the simulations conducted in the present study. From Fig. 1 of "A PDP Model of the Simultaneous Perception of Multiple Objects," by C. M. Henderson and J. L. McClelland, 2011, *Connection Science*, *23*, p. 163. Copyright 2011 by Taylor & Francis. Adapted with permission

to identify a particular configuration of learning rate, number of training epochs, or initial weights that would optimize the match between ANN performance and behavioral performance).

Summary of the Church et al. (2010) study Church et al. (2010) compared the perceptual category-learning and generalization abilities of children with HFASD (primarily Asperger's disorder) to those of typically developing children who were matched with respect to age (range 7–12), IQ ($M = 110$; $SD = 10$), and demographic characteristics. In the training phase of the categorization task (consisting of 30 trials), the children were introduced to shapes that they were told would or would not belong to the category "cave ghosts," and that they would be expected to classify during the testing phase. These two-dimensional shapes were constructed from patterns of dots comparable to those described by Posner and colleagues (Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968). The visual images used in the study were members of one of two classes: (1) shapes that were systematically distorted versions of an original "canonical" prototype shape (we refer to this class of shapes as *prototype distortions*, or simply *distortions*), corresponding to the "cave ghost" class of stimuli; or (2) random shapes that had no relationship to the prototype, and therefore did not belong to the class "cave ghost." The canonical prototype was created by randomly generating a set of dots within a fixed area, connecting those dots to create polygons, and then filling the polygons with a random color. Distorted versions of the prototype were created by varying the probability that each dot would move from its original position. The prototype distortions varied by level of distortion (L), with higher levels indicating less resemblance to the canonical prototype.[1] To generate the random class of stimuli, dots randomly placed within a fixed area were connected and then filled with a random color. The random shapes bore no meaningful resemblance to each other, nor did they resemble the distortions. Figure 2 presents examples of the stimuli used in the Church et al. study.

Individual shapes were presented on the computer screen in random order, and participants pressed a button to indicate whether the image was or was not a "cave ghost." This task has been described in the category-learning literature as an (A, not A) task (Ashby & Maddox, 2005). Each shape appeared only once during training. During the training phase, feedback was presented onscreen to indicate whether the response was correct. The participants were

[1] Due to the probabilistic method by which stimuli were generated, the similarity between the canonical prototype and a particular stimulus may not be representative of the group's average similarity to the canonical prototype. For a full description of the method used to generate the stimuli, refer to the electronic supplement of Church et al. (2010).
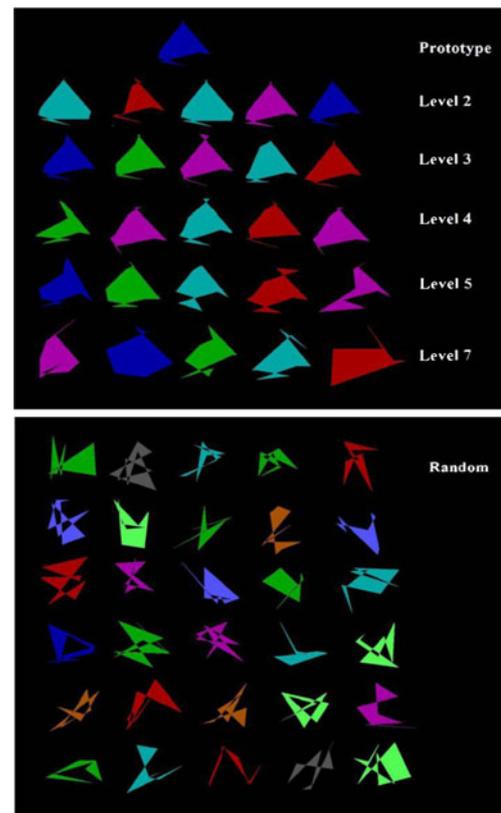
Fig. 2 Examples of visual stimuli used by Church et al. (2010). Children were trained to classify shapes from Levels 3, 5, and 7 as "cave ghosts," and to classify random shapes as not being cave ghosts. Lower levels correspond to less distorted versions of the prototype, and therefore are more similar to each other and to the prototype

subsequently tested with additional category members using the same procedure as during training, with the exception that participants did not receive feedback during the testing phase. Figure 3 shows the proportions of novel shapes that children in both groups classified as being cave ghosts.

Network inputs and outputs The visual stimuli created by Church et al. (2010) provided the basis for the input set. The training images included 15 prototype distortions (five L3, L5, and L7 distortions), as well as 15 random shapes. The test images consisted of the canonical prototype (L0) and 25 distortions (five each of the L2, L3, L4, L5, and L7 shapes), as well as 30 random shapes. The original images from Church et al.'s training and test sets were converted into matrices representing features within the images (see the supplemental materials for details).

The target outputs in our simulations included only two possible outputs corresponding to right and left button-presses. In contrast, Henderson and McClelland's (2011) model was originally designed to associate objects at different locations with various actions performed in ways that depended on the location of the object, and thus involved
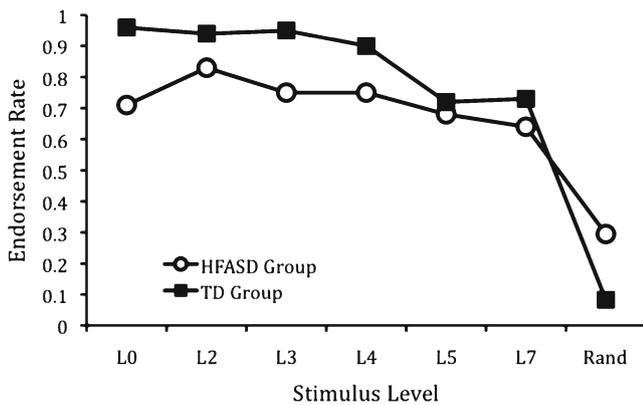
Fig. 3 Church et al.'s (2010) behavioral results. The endorsement rates for the typically developing (TD) children and children with high-functioning autism spectrum disorder (HFASD) are shown for each level of prototype distortion (L0 is the prototype, and L2–L7 denote increasing levels of distortion), as well as for the random shapes. Overall, children with HFASD were more likely to endorse novel low-level distortions of a prototype than the actual prototype, and were generally less likely to endorse novel prototype distortions

more variable and complex target outputs than those used in the present simulations. Our two outputs were encoded using a 144-element array to match the output layer size of the original model.

Results and discussion

We initially verified that Henderson and McClelland's (2011) dorsal-only ANN could learn to classify inputs as either prototype distortions or random shapes. The results reported below reflect the averages of 20 distinct simulations with randomly initialized networks; the number of ANNs tested per condition corresponded to the number of typically developing children in the Church et al. (2010) study.

Given an appropriate learning rate and a sufficient number of training epochs, the networks achieved almost perfect categorization accuracy. Because of the relatively small amount of training in Church et al.'s (2010) original behavioral experiment (30 trials total), few children attained such high performance levels. We found that training the ANNs for 15 epochs at a learning rate of .0001 produced a close approximation to the TD group's generalization pattern (Fig. 4a).[2] This ANN configuration (hereafter referred to as the *TD model*) provided the basis for subsequent manipulations aimed at simulating perceptual generalization by children with HFASD. There was relatively little variation in the generalization performance of the TD models. As noted earlier, other existing computational models developed

_____
[2] These parameter settings were chosen somewhat arbitrarily; similar generalization patterns would likely be produced by ANNs with other settings.

specifically to model perceptual category learning would likely be capable of generating similar generalization patterns.

The abstract encodings of images used to train ANNs in this study did not match the inputs that Henderson and McClelland's (2011) model of visual processing was originally designed to classify, and undoubtedly differed in many respects from how the children's brains encoded those same images. Consequently, the fact that the ANNs classified novel shapes in ways similar to the TD children suggests that this model captures general properties of visual category learning and generalization, and that the observed generalization patterns are not simply an idiosyncratic feature of the inputs or parameters selected for these simulations. These findings establish that a generic computational model of visual object processing could learn and generalize visual shape classifications in ways that led to endorsement rates comparable to those produced by typically developing children.

Simulation 2

Simulating visual category learning by children with HFASD

Church et al. (2010) found that children with HFASD were less likely than typically developing children to endorse prototype class stimuli (especially the prototype stimulus) and more prone to endorse random stimuli as members of the category. Given that the canonical prototype is the shape most representative of the characteristics of the prototype-based family of images in this task, TD participants usually classify it as being a member of the learned category (Posner & Keele, 1968; see also Fig. 3 above). The fact that children with HFASD were less likely to endorse the prototype than some distortions suggests that they did not simply perform more poorly than TD children, but showed a qualitatively different pattern of learning and generalization.

We used four different manipulations in an attempt to simulate this atypical generalization pattern, including decreasing the rate of learning, disrupting changes in connections during learning, modifying the number of processing units within the ANNs, and decreasing the effective signal-to-noise ratio in the processing units. Below we describe how each of these manipulations was computationally instantiated, as well as the efficacy of each approach. We also conducted additional simulations to generate specific behavioral predictions that could further differentiate the potential impacts of these different manipulations in novel training conditions.

Description of the model

Architecture and implementation The ANNs used in these simulations, as well as the inputs and outputs, were identical

Fig. 4 (**a**) Comparison of the typically developing (TD) model and TD group endorsement rates for the test stimuli. The endorsement rates show that the model is able to accurately simulate the Church et al. (2010) TD perceptual generalization results for all stimulus distortion levels. (**b**) Decreasing the learning rate produced a reasonable approximation of generalization by children with HFASD, with the exception that the neural networks were more likely to endorse L0, the canonical prototype. (**c**) The negative-weight-decay manipulation also produced a reasonable approximation, aside from the prototype endorsement rate. (**d**) Increasing hidden-layer size did not replicate atypical generalization by individuals with HFASD, and instead improved the network's endorsement accuracy. (**e**) Changes in gain values also did not replicate the observed atypical generalization
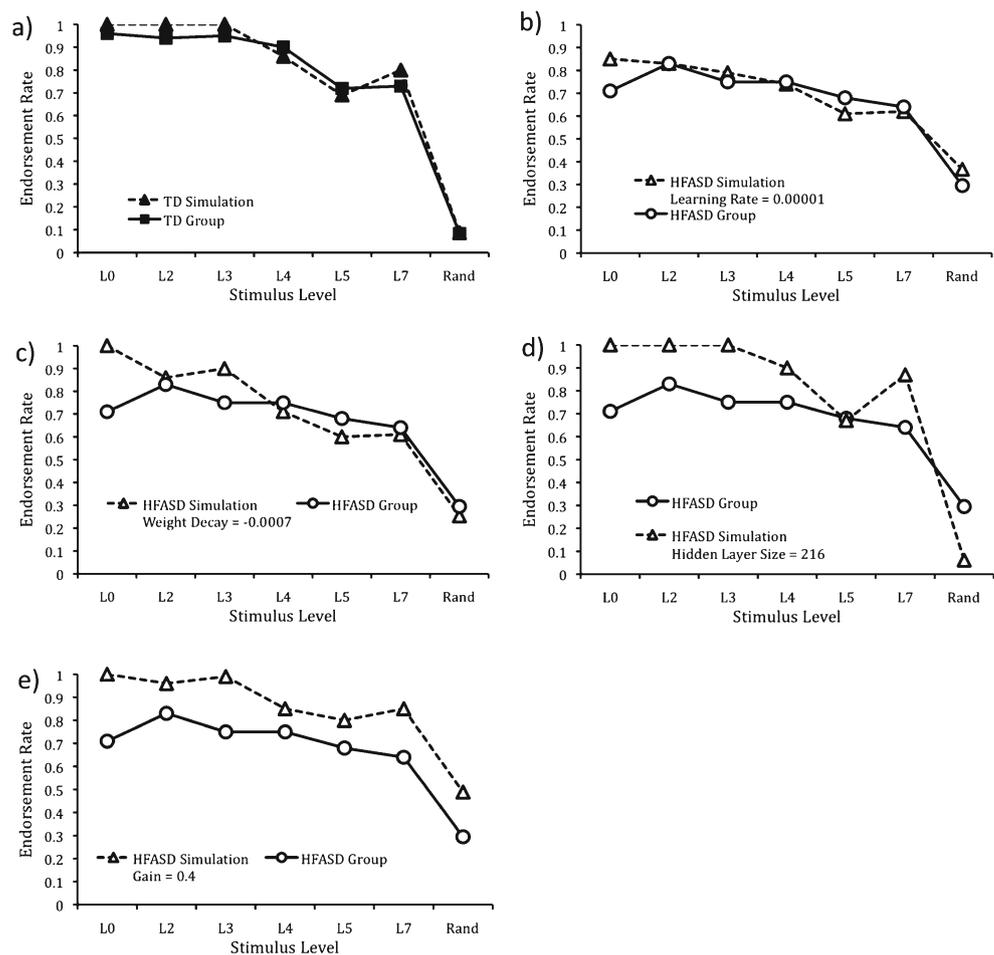
to those in Simulation 1. We simulated deficits in neural plasticity and neural homeostasis associated with HFASD by adjusting the learning rates and weight decay settings of the ANNs. Just as structural and functional synaptic abnormalities can impede the learning process, changes to a neural network's learning rate or weight decay values can hinder the network's learning and generalization. We simulated the cortical structural abnormalities associated with HFASD by increasing and decreasing the number of hidden units in the ANNs, and also simulated increased neural noise by adjusting the gain of the activation functions within these hidden units. No attempt was made to search parameter space for optimally fitting parameters. We initially tested a broad range of parameter settings, after which we tested a smaller range surrounding the parameter values that led to generalization most similar to that of children with HFASD.

Simulating degraded plasticity The learning rate (LR) parameter of an ANN scales the magnitude of the possible changes made to network weights during training. If a network's learning rate is too low, the ANN might take an excessive amount of time to reach an adequate level of performance. A large number of training epochs combined

with a low learning rate might also cause the network to *overfit* the data, leading to poor generalization (Dawson, 2005). We attempted to simulate atypical generalization by children with HFASD by lowering the TD model's learning rate. This manipulation may be considered analogous to reducing synaptic plasticity in biological neural systems. The overall effect of this manipulation was to decrease the rate at which task-relevant learning occurred.

Simulating degraded homeostasis The weight decay parameter of an ANN can facilitate generalization through the adjustment of the network's error function, decreasing the variability of the weight changes during learning (Bishop, 1995); this technique, and other methods that potentially discourage overfitting, are known as "regularization" methods. We used an "anti-regularization" method (Hamamoto, Mitani, Hase, & Tomita, 1997; Raudys, 1998, 2001) implemented by adding a term scaled by the weight decay value to the network's learning algorithm (potentially encouraging overfitting), to attempt to replicate the HFASD generalization data. The effect of this manipulation was to distort the magnitude of changes in the network's weights during each cycle of the training process. This manipulation of the

ANNs can be viewed as analogous to altering the ability of neural circuits to establish homeostasis during learning.

Simulating structural abnormalities Under certain conditions, increasing the number of units in an ANN's hidden layer can decrease its ability to correctly categorize unfamiliar stimuli (Cohen, 1994; Murata, Yoshizawa, & Amari, 1994). In other words, an excessively large hidden layer can compromise a network's ability to generalize and can lead to overfitting of the training set, which could potentially produce effects comparable to the impaired generalization observed in individuals with autism (Cohen, 1994). In other cases, however, a large hidden layer can improve recognition *and* generalization (Caruana, Lawrence, & Giles, 2001; Cohen, 1998). Drawing a possible parallel between the effects of an excessive number of hidden-layer units in ANNs and increased minicolumn density in people with ASD, we examined whether an increase or decrease in the number of hidden-layer units would produce atypical generalization.

Simulating increased neural noise A decrease in signal-to-noise ratio can be modeled in a neural network by reducing the slope of the model's activation function, referred to as *lowering the gain* (Fellous & Linster, 1998). Gain modulation has neurological correlates in catecholamine-modulated signal-to-noise ratio changes (Servan-Schreiber, Printz, & Cohen, 1990) and could also correspond to changes in cortical excitability resulting from disruptions in GABAergic (Rubenstein & Merzenich, 2003) or GABAergic and glutaminergic (Vattikuti & Chow, 2010) systems, as has been proposed to occur in individuals with autism. We attempted to simulate atypical generalization associated with a decreased signal-to-noise ratio by lowering the TD model's gain. Because the PDPTool software does not provide direct access to the properties of hidden-unit transfer functions, gain reductions were instead implemented in accordance with the method specified by Li, von Oertzen, and Lindenberger (2006): Initial weights were scaled by the gain value, and a new learning rate was used that was equal to the original learning rate multiplied by the square of the gain value. Noisy stimulus representations could result from an insufficiency or overabundance of neurotransmitters or neuromodulators in neural circuits, among other factors, and might, in principle, account for degraded recognition of the prototype.

Predicting generalization in novel situations Training with different input sets may lead to different patterns of generalization across various methods of simulating HFASD performance. Simulations of different training regimens thus provide a way to examine how well different manipulations predict category-learning and generalization differences across groups. To explore this possibility further, we trained ANNs with a novel input set, consisting of the L0 prototype as the sole exemplar of a cave ghost, as well as all of the random stimuli included in the original training set.

Results and discussion

Table 1 summarizes how parameters were manipulated in these simulations, the neural abnormalities that these manipulations were designed to simulate, and the capacity of these manipulations to account for atypical generalization in children with HFASD.

Modeling atypical generalization A reasonable match to the HFASD generalization pattern was found when LR was decreased to .00001—an order of magnitude lower than the learning rate used in the TD model (Fig. 4b). Decreasing LR reduced the likelihood of endorsement at all prototype distortion levels (i.e., there was a global drop in performance). Consistent with our simulation results, multiple reports have indicated that individuals with autism often take longer than TD participants to learn to categorize visual items (Bott et al., 2006; Schipul, 2012; Soulières et al., 2011; Vladusich et al., 2010). Also, ASD is associated with atypical synaptogenesis and synaptic function (Auerbach et al., 2011), which could disrupt synaptic plasticity and, in turn, degrade learning mechanisms (Wang et al., 2011). Although decreasing LR led to worse generalization relative to the TD model, this manipulation did not lead to lower endorsement of the prototype (L0) relative to distortions, in contrast to what was observed behaviorally. This discrepancy suggests that although the reduced-LR model captures some important aspects of perceptual category learning by individuals with autism, it does not tell the whole story.

A similar shift in generalization was also found when negative weight decay was added to the TD model (Fig. 4c). Although this manipulation provided a reasonable fit to the HFASD results, models with negative weight decay again endorsed the prototype more often than did children with HFASD. An interesting feature of networks with negative weight decay was that they showed significant sensitivity to the network's initial weights, such that networks with the same parameter values could produce dramatically different generalization patterns. The high level of individual variability in these simulations thus captures the significant response variability observed when people with ASD perform perceptual and cognitive tasks (Mottron et al., 2006). Learning abnormalities characteristic of the autistic phenotype have been associated with a diminished ability to establish neural homeostasis (Bourgeron, 2009; Ramocki & Zoghbi, 2008), which may lead to analogous disruptions of synaptic plasticity.

Increasing the number of hidden units did not result in generalization patterns comparable to those observed in the HFASD group (Fig. 4d). On the contrary, this manipulation maintained or slightly improved the ANNs' ability to correctly

Table 1 Summary of neural-network parameter values for the simulation results shown in Fig. 4 and Fig. 8, along with the atypical neural process that each manipulation is intended to emulate

Model Variant	Neural Abnormality Simulated	Ability to Account for Atypical Generalization	Parameter				
			Number of Training Epochs	Learning Rate	Momentum	Weight Decay	Hidden-Layer Size
TD	N/A	N/A	15	.0001	.9	0	144
ASD-LR	Reduced neural plasticity	High	15	.00001	.9	0	144
ASD-WD	Reduced neural homeostasis	High	15	.0001	.9	−.0007	144
ASD-gain	Increased neural noise	Low	15	.000016	.9	0	144
ASD-hidden layer	Increased number of minicolumns	Low	15	.0001	.9	0	216
ASD-LR *A Type II*	Reduced neural plasticity	High	15	.000005	.9	0	144

TD, typical development model; ASD, autism spectrum disorder; LR, reduced-learning-rate model; WD, weight decay model; gain, lowered-gain model; hidden layer, expanded-hidden-layer model; LR *A Type II*, reduced-learning-rate model for A Type II subgroup.

endorse prototypes and diminished the number of random stimuli that were endorsed. Decreasing the number of hidden units below a certain threshold degraded generalization, but not in ways that were comparable to children with HFASD. These findings are in contrast with earlier simulations in which increasing the size of the hidden layer within the model did degrade generalization (Cohen, 1994, 1998).

We also did not find a network gain setting that could approximate the HFASD group's generalization pattern. Figure 4e shows the model's performance when gain was 40 % of the TD model's original gain, producing one of the closest approximations to the atypical generalization observed in children with HFASD. Thus, we did not find evidence to support the idea that an atypical signal-to-noise ratio is responsible for atypical generalization. It is important to note, however, that the gain manipulation only degraded the signal-to-noise ratio during acquisition, and therefore did not account for possible effects of degraded processing during testing.

Quantitative comparisons of mean differences in generalization patterns from those observed behaviorally for each of the four ANN manipulations confirmed that decreasing LR provided the best fit ($M = .05$, $SD = .04$), followed by adding negative weight decay ($M = .09$, $SD = .09$), decreasing gain ($M = .18$, $SD = .07$), and increasing the number of hidden units ($M = .19$, $SD = .09$). In all simulations, the largest difference from the observed generalization performance was that the ANNs were more likely to endorse the prototype ($M = .22$, $SD = .09$).

Predicting generalization Figure 5a shows how the TD model generalized after training with the new input set (contrasted with the results of Simulation 1). These simulations predict that after training with prototypes and random stimuli, TD children would be adept at endorsing stimuli at low distortion levels, but less likely to endorse highly distorted stimuli than they were in the original study.

Of the manipulations described above, decreasing learning rate or adding negative weight decay to ANNs produced the closest approximations to the HFASD generalization pattern. We anticipated that these two manipulations would produce distinctly different patterns of generalization when the ANNs were trained to distinguish prototypes from random stimuli. When trained with the new input set, HFASD simulations using a reduced LR showed a smaller decline in endorsement rates with higher levels of distortion than were predicted by the TD model (Fig. 5b), and they made the strong prediction that children with HFASD should be *more likely* than TD children to endorse highly distorted prototypes as members of the learned category.

After training with the new input set, ANNs with negative weight decay showed substantially lower endorsements of L2 and L3 shapes than either the reduced-LR or TD neural networks (Fig. 5c), and were also less likely to endorse other distortions. This manipulation thus predicts that if category learning by children with HFASD is negatively impacted by dysfunctional synaptic homeostasis, then children with HFASD should show worse generalization after training with the novel input set. This prediction is compatible with previous suggestions that autism is associated with hyperspecific learning (Grossberg & Seidman, 2006; Markram & Markram, 2010; McClelland, 2000), because hyperspecific learning should lead to steep generalization gradients.

Simulation 3

Simulating visual category learning by subgroups of children with HFASD

None of the parameter manipulations tested above proved to be able to fully account for the atypical pattern of
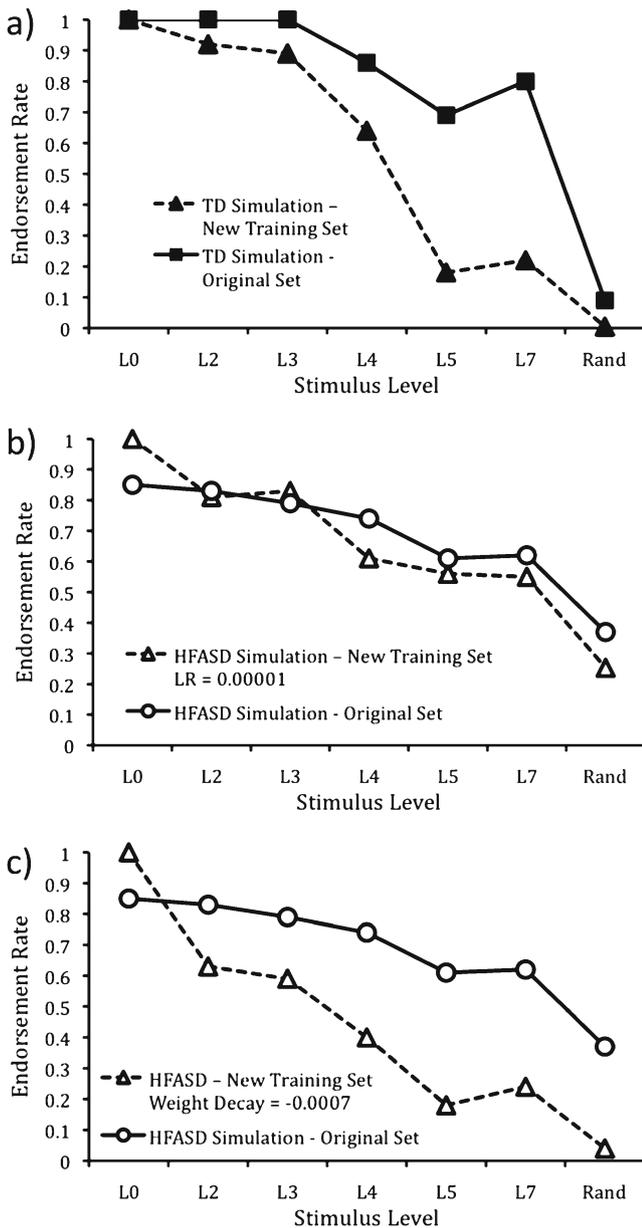
Fig. 5 (a) The model predicts that TD participants trained only with prototypes would demonstrate low accuracy in endorsing highly distorted prototypes. (b) Reduced-learning-rate simulations predict that children with HFASD trained only with the prototype would show a moderate decline in endorsement accuracy at higher levels of distortion. (c) Simulations based on manipulating negative weight decay predict that children with HFASD trained only with the prototype would show a considerable decline in endorsement rates at all stimulus levels above L0, with the highest distortion levels showing the largest declines

generalization observed by Church et al. (2010). This discrepancy could mean that the model or model parameters failed to capture some key process underlying differences in visual category learning and generalization across groups. Alternatively, children with HFASD may show a heterogeneous pattern of learning and generalization, such that the pattern evident at the group level is not representative of the

patterns shown by individual children. To explore this possibility, the generalization patterns produced by individual children in the Church et al. study were automatically sorted using a self-organizing map (SOM), and then simulations comparable to those above were run in an attempt to replicate the generalization patterns shown by the subgroups of children identified by the SOM.

Description of the model

Architecture and implementation The ANNs used in these simulations, as well as their inputs and outputs, were again identical to those used in Simulation 1. Generalization profiles for individual children were generated from test performance (seven measures for each child; $n = 40$), and then used to train an SOM created with the data-mining program Orange (Demsar, Zupan, Leban, & Curk, 2004). SOMs are a type of neural network in which the spatial arrangement of units within the network becomes organized through training, such that adjacent nodes respond to similar inputs (Kohonen, 2001). The prevalence of repeated input features determines the response properties of each node after training. Nodes acquire their selectivity to features by competing to match each input during training; the sensitivities of the winner node and its neighbors are gradually adjusted to increase their responsiveness to "matching" inputs. The resulting map thus provides a way to identify the prevalence of different input properties (in this case, different patterns of generalization), as well as how those properties vary across individuals (Mercado, 2011).

Results and discussion

Figure 6 shows the output of a 4 × 5 SOM that was trained with the 40 generalization profiles from children in the
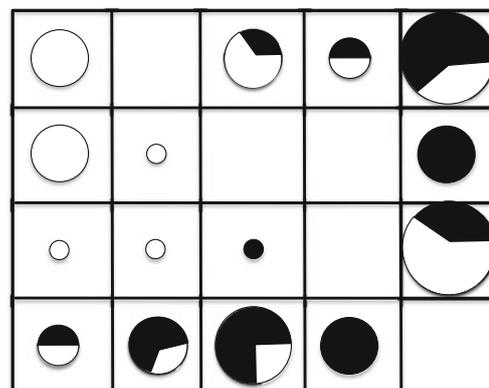
Fig. 6 Output of a 4 × 5 self-organizing map (SOM) trained with generalization profiles from Church et al. (2010). Circle size indicates how many children best matched that unit in the trained map (range = 1–5). Typically developing children are shown in black, and HFASD children in white

Church et al. ([2010]) study. Each square corresponds to a processing unit in the SOM, and the size of the circle within each square indicates the number of children whose profile best matches a particular unit (the smallest circles correspond to one child, and the largest to five children). To reveal the structure of the map, circles were color-coded according to whether the children were typically developing (black) or had HFASD (white). The pie charts show the proportions of children from each of these two groups associated with each map unit.

Analyses of spatially contiguous clusters of nodes within the SOM were used to calculate prototypical generalization profiles for subgroups of similar individuals. These analyses revealed two subsets of children with HFASD, one associated with the upper left corner of the map, and the other associated with the right edge of the map. One subgroup of children with HFASD (hereafter referred to as *A Type I*; *n* = 11) showed a generalization pattern that was essentially indistinguishable from that of TD children, whereas the other subgroup (referred to as *A Type II*; *n* = 9) showed little generalization. Examination of the generalization patterns revealed that children within each subgroup could be identified on the basis of the rates at which they endorsed random stimuli as cave ghosts during testing. Children who endorsed more than 30 % of the random stimuli as cave ghosts showed an A Type II profile, whereas children who endorsed fewer random shapes showed an A Type I profile (Fig. [7a]).

Initial attempts to produce a pattern of generalization comparable to A Type II children by either decreasing the learning rate or increasing negative weight decay proved unsuccessful. In all cases, ANNs endorsed the prototype more than other stimuli (as in Simulation 2). A subset of ANNs did produce generalization patterns similar to the A Type II endorsement pattern. In particular, by averaging across the generalization patterns of nine reduced-learning-rate (LR = .000005) ANNs that best matched the performance of individual children with A Type II profiles (selected from 100 simulations), we were able to construct a generalization profile that closely matched their endorsement pattern (Fig. [7b]). Thus, reduced-LR ANNs are capable of generalizing like children with A Type II profiles but are too variable in performance to make precise predictions about how small groups of children with A Type II profiles will generalize. In part, this is because with few training trials and a low learning rate, the initial weight settings (which are randomized) will have a larger impact on how an ANN generalizes.

Overall, these findings provide an explanation for why the ANNs in Simulation 2 were unable to recreate the overall patterns of prototype endorsement by children with HFASD. The group-level models in Simulation 2 assumed
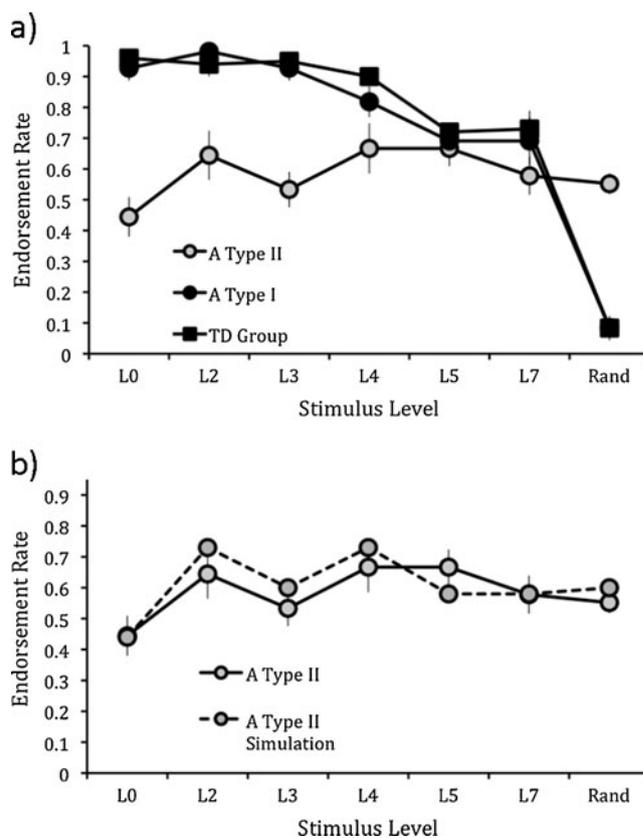
Fig. 7 (**a**) Generalization profiles for two subgroups of children with HFASD, identified by the SOM shown in Fig. 6. Children who endorsed fewer than 30 % of the random stimuli as cave ghosts (*n* = 11) showed generalization comparable to the TD children (A Type I), whereas children who endorsed more random shapes (*n* = 9) showed little generalization (A Type II). (**b**) Simulations in which individual reduced-learning-rate ANNs were matched to individual generalization profiles produced an overall generalization pattern comparable to that of the A Type II profile

that generalization in children with HFASD could be characterized in terms of systematic deviations in learning or stimulus-processing mechanisms. Instead, the analyses above suggest that a subset of individuals with HFASD were successful at the visual category-learning task, whereas others were not. No uniform adjustment to the model parameters would be able to capture such dichotomous outcomes. Instead, the model must take into account individual differences in learning to fully characterize the empirically observed patterns (see also Lee & Webb, [2005]; Nosofsky, Palmeri, & McKinley, [1994]; Smith, Murray, & Minda, [1997]). It is the linear combination of these two dichotomous generalization profiles (associated with the two subgroups of children with HFASD shown in Fig. [7a]) that leads to the lowering of prototype endorsement by children with HFASD. Aggregation of heterogeneous subgroups of individuals with HFASD may also lead to seemingly contradictory findings in studies of adults, as is illustrated by Simulation 4.

Simulation 4

Simulating the Vladusich et al. (2010) experiments

Only one study other than Church et al. (2010) has directly measured the endorsement of novel dot patterns after feedback-based training in individuals with HFASD. In two experiments, Vladusich et al. (2010) trained typical adults and adults with HFASD on an (A, B) visual category-learning task in which they had to classify dot patterns with medium levels of distortion into two categories. The researchers then tested participants with sets of novel and familiar dot patterns of various distortion levels. In both experiments, Vladusich et al. reported that individuals with HFASD showed a typical pattern of generalization, indicating intact prototype recognition. However, participants were trained to a fixed criterion before being tested, and in at least one of the two experiments, individuals with HFASD proved to be much slower at learning the task. The following simulations assessed whether the simple model of visual object perception described above could account for the apparent discrepancies between Church et al.'s finding that visual category learning is disrupted in children with HFASD and Vladusich et al.'s finding that category learning in adults with HFASD is intact.

Description of the model

Architecture and implementation Initial attempts to simulate the (A, B) visual category-learning task with a dorsal-only model proved to be unsuccessful. In all cases, the trained ANNs were more likely to endorse dot patterns with medium levels of distortion than either prototypes or dot patterns with low-level distortion. This outcome is consistent with past suggestions that (A, B) category learning engages learning mechanisms beyond those required for an (A, not A) category-learning task (Ashby & Maddox, 2005). We thus switched to testing with the ventral-only model (described in Simulation 1), which was specifically designed by Henderson and McClelland (2011) to flexibly categorize objects. This model differs from the dorsal-only model in that it has two hidden layers and recurrent connections within the second hidden layer (see Fig. 1). The additional hidden layer and internal feedback within the ventral-only model increases its capacity to represent idiosyncratic information that identifies particular categories. The architectural specifications and parameters used in this model were identical to those used by Henderson and McClelland, with the exceptions that the learning rates and target outputs were modified to match those of Simulations 1 and 2. ANNs with a learning rate of .0001 were used to simulate performance by typical adults (as in Simulation 1), as well as adults with HFASD who learned at rates comparable to those of typical

adults. On the basis of the results of Simulation 3, which showed that the performance of children with an A Type I profile could be simulated using the same learning rate that was used to simulate performance by typical children, we classified adults with a typical learning rate as having an A Type I profile. Adults with HFASD that showed slower than typical learning rates were classified as having an A Type II profile, based on the results of Simulation 3 that showed that the performance of children with an A Type II profile could be reproduced by lowering learning rate. These slower-learning adults were simulated using ANNs with a reduced learning rate (.00001). The ANNs used to simulate performance by typical adults (TA) were trained to a criterion of 80 % or more correct classifications, following the experimental design of Vladusich et al. (2010). The ANNs used to simulate performance by HFASD adults with an A Type II profile were trained until they had completed three times as many epochs as the average amount required by TA models to achieve criterion (again matching the effective protocol of Vladusich et al., 2010). The TA simulations required an average of 39 epochs to train to criterion. All of the A Type II networks were trained for 117 epochs. Ten ANNs were trained for each condition, and performance was averaged across the ten simulations. Overall performance of the simulated HFASD group was calculated using a weighted average of the performance of ANNs with A Type I profiles and those with A Type II profiles, as described below.

Inputs and outputs The specific stimuli used by Vladusich et al. (2010) are not publicly available, and even if they were, the image-coding scheme used in Simulations 1–3 was not appropriate for describing dot patterns in which the dots are not connected. Consequently, we developed a new set of 144-element input vectors based on the small set of sample stimuli reported by Vladusich et al. Matrices corresponding to the two reported prototypical dot patterns were created, and then distorted versions of these patterns were generated that had low, medium, and high levels of distortion. All ANNs were trained with a single set of stimuli made up of 32 medium-level distortions. After training, the networks were tested on their responses to the 32 training stimuli (in Fig. 8, MF = medium, familiar); 32 medium-level-distortion items that were novel (MN = medium, novel); 32 low-level-distortion items that were novel (LN = low, novel); and 32 high-level-distortion items that were novel (HN = high, novel). This distribution of testing and training stimuli paralleled the distribution of stimuli used by Vladusich et al.

Results and discussion

Simulations of category learning by typical adults using the ventral-only model produced a generalization pattern comparable to that reported by Vladusich et al. (2010) for all
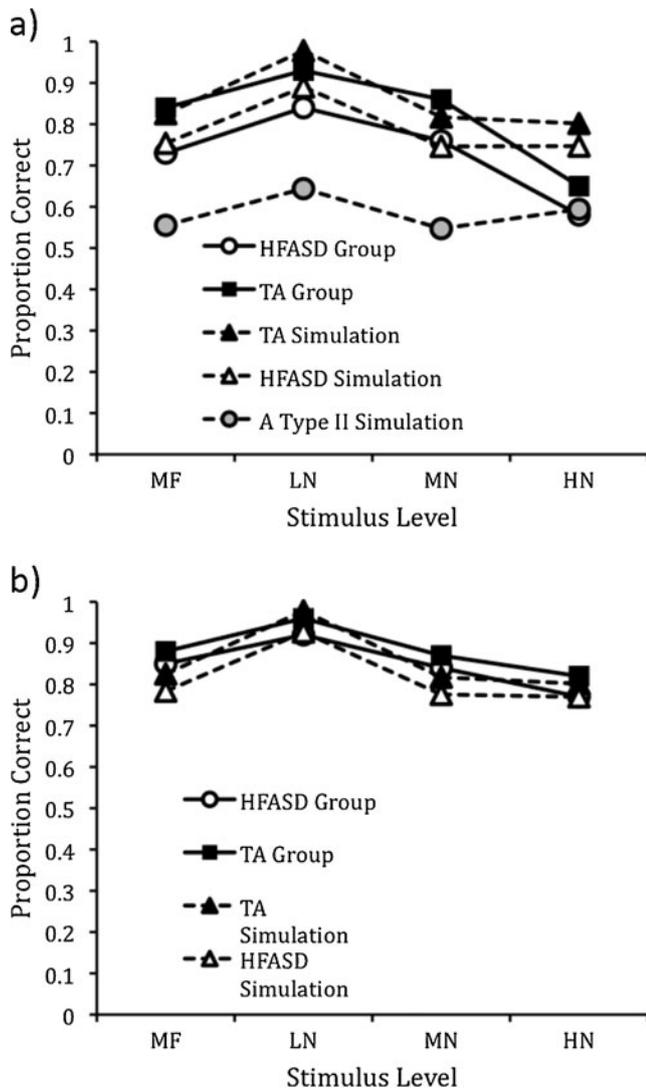
Fig. 8 (a) Vladusich et al.'s (2010) Experiment 1 showed similar generalization profiles for both typical adults (TA) and adults with HFASD after the participants were trained to identify two different categories. The trained stimuli were medium-level distortions of two different dot patterns (MF), and the novel stimuli were dot patterns with low (LN), medium (MN), or high (HN) levels of distortion. Simulations with Henderson and McClelland's (2011) ventral-only model generalized similarly to TA groups for all of the stimulus classes except high distortions. Reduced-learning-rate simulations of generalization by adults with HFASD in which the participants were divided into those with A Type I or A Type II profiles showed an overall drop in generalization performance comparable to that reported by Vladusich et al. (b) Vladusich et al.'s Experiment 2 showed nearly identical generalization profiles for both typical adults and adults with HFASD. The stimuli here were distorted dot patterns generated using a different procedure from that of Experiment 1. Reduced-learning-rate simulations of generalization by adults with HFASD in which the participants were divided into those with A Type I and A Type II profiles showed the same small drop in generalization performance observed experimentally

stimuli other than the high-level-distortion stimuli (Fig. 8a). The A Type II simulations produced a generalization pattern

much worse than was reported for their adults with HFASD (Fig. 8a). However, as was illustrated by Simulation 3, not all individuals with HFASD show an A Type II generalization pattern. Vladusich et al. reported individual learning curves and explicitly noted in discussing the results of Experiment 1 that several individuals with HFASD took much longer than was typical to reach criterion during training. Figure 2 from their article shows that five individuals with HFASD were trained for more blocks than any typical adult. If these five individuals are assumed to show A Type II profiles and the remaining 14 participants are assumed to show A Type I profiles (i.e., generalization comparable to a typical adult), then the overall pattern of generalization by adults with HFASD can be compared to a similarly weighted average of ANN performance. This averaged generalization pattern shows an overall drop in performance across all stimuli, consistent with the findings of Vladusich et al. (Fig. 8a).

Vladusich et al. (2010) performed a second experiment that was identical to their first, except that they used different methods to generate prototype distortions. In their second experiment, they found no statistically significant differences in generalization performance between the groups (Fig. 8b). Vladusich et al. attributed the better performance of individuals with HFASD in this second experiment to the differences in stimulus construction. The present simulations cannot replicate the differences in the images used across these two experiments. However, the learning profiles reported by Vladusich et al. in Fig. 4 of their report show that in their second experiment, only two of 13 individuals with HFASD required more blocks of training than any typical adult. If it is assumed that these two participants would fall within the A Type II subgroup, and that the remaining 11 adults with HFASD showed A Type I profiles, then a weighted average of the generalization patterns for this distribution of individuals reveals generalization comparable to that reported by Vladusich et al. (Fig. 8b). Thus, the ventral-only model can account for the lack of statistically significant differences that they observed in Experiment 2 without needing to assume that individuals with HFASD were better able to process the novel stimulus sets. Specifically, the present model predicts that when a sample of adults (or children) with HFASD is skewed toward individuals with A Type I profiles, the atypical generalization patterns of individuals with A Type II profiles are unlikely to be statistically detectable in between-group comparisons because of the large within-group variance, and because A Type I profiles are essentially indistinguishable from typical generalization patterns.

It is important to note that the same changes in learning rate that adequately re-created generalization differences observed in children with HFASD by Church et al. (2010) when used with dorsal-only ANNs also proved to be able to

account for the differences reported by Vladusich et al. (2010) in adults with HFASD when applied to ventral-only ANNs. This correspondence across network architectures is consistent with the proposal that basic learning mechanisms are disrupted in individuals with HFASD because such deficits should affect processing throughout cortical networks. In other words, learning an (A, B) visual categorization task may engage neural systems that are not required when learning an (A, not A) categorization task, but if neural plasticity is globally disrupted, then both systems are likely to show the impacts of such a deficit, even if the specific effects of reduced plasticity on learning and generalization differ across these systems.

General discussion

In this study, we used neural network models to explore how cortical deficits may contribute to variations in category learning and generalization by individuals with HFASD. We discovered that subcomponents of Henderson and McClelland's (2011) model of visual object perception were sufficient to simulate generalization after training on a visual category-learning task. These simulations were guided by a recently proposed framework for understanding cortical constraints on learning capacity (Mercado, 2008, 2011), as well as recent hypotheses about how neural abnormalities may contribute to the cognitive symptoms of autism (Markram & Markram, 2010; Rubenstein & Merzenich, 2003). The simulations suggest a simple, yet intriguing explanation for why several past studies of category learning by individuals with autism have produced seemingly contradictory results. Specifically, the analyses here (especially Simulation 3) suggest that some individuals with HFASD show dramatic impairments when learning to classify abstract visual patterns, whereas others show no impairment. In the following sections, we consider why there are such large disparities in perceptual category learning and generalization across individuals with HFASD, and how computational models can clarify the contributions of neural abnormalities to such disparities.

Individual variations in category learning

Typically developing children can vary considerably in learning capacity. Nevertheless, most of the TD children examined by Church et al. (2010) learned to identify "cave ghosts" at near-ceiling levels. The A Type I cluster of children with HFASD also rapidly learned to classify shapes, confirming several past reports that individuals with HFASD have the capacity to learn visual categories and to recognize prototypes as members of those categories (e.g., Bott et al., 2006; Soulières et al., 2011). Vladusich et al.

(2010) reported that when adults with HFASD were trained and tested on an easy visual categorization task, all of the individuals with HFASD generalized as did typical adults. When these same adults were trained and tested on a more difficult task, however, differences in learning and generalization emerged. Gastgeb et al. (2012) noted that some adults with HFASD rapidly learned visual categories when trained without feedback, whereas others did not. They attributed these differences to variations in nonverbal IQ, suggesting that individuals with a higher nonverbal IQ were better able to discover alternative strategies for learning the task. In contrast, other researchers have suggested that when individuals with ASD show deficits in category learning, these deficits are a side effect of other degraded capacities. For example, Molesworth et al. (2008) proposed that individuals who failed to recognize prototypes in generalization tasks might do so because they failed to understand the task demands or because of a lower mental age. Explanations of atypical generalization based on variations in mental age or IQ are unlikely to account for the differences reported by Church et al., because those children were all high functioning, and because differences in IQ across participants were small and matched to those of TD children.

Heterogeneity in the capacities of individuals with HFASD is not specific to category learning and is evident across social impairments as well as physiological responses (Hirstein, Iversen, & Ramachandran, 2001). The present simulations (especially Simulation 4) suggest that experimental studies of category learning that do not account for possible dichotomous differences in subgroups of individuals with HFASD might show either no differences from typical categorization performance or large differences from typical performance, depending on the mixture of A Type I and A Type II profiles present within groups of individuals with autism. This factor may account for many of the discrepancies in past reports about whether individuals with HFASD are impaired at learning perceptual categories. Several researchers have suggested that individuals with HFASD can overcome deficits in category learning with sufficient training experience, ultimately showing typical levels of performance (Bott et al., 2006; Schipul, 2012; Soulières et al., 2011; Vladusich et al., 2010). The present simulations are consistent with these proposals, in that reduced-LR ANNs can eventually learn to classify shapes appropriately.

It remains unclear how individuals showing an A Type II profile differ from those with A Type I profiles. Atypical learning and generalization by individuals with HFASD might reflect differences in how novel two-dimensional shapes are encoded (Samson, Mottron, Soulières, & Zeffiro, 2012; Sheppard, Ropar, & Mitchell, 2009) or differences in how abstract classification problems are approached (Soulières et al., 2011). Differences in performance may also

reflect intrinsic differences between the two subgroups. Regardless of the source(s) of these differences, the present simulations show that atypical generalization by individuals with HFASD can be accounted for by changing a single parameter (learning rate) within a relatively simple ANN model of visual object perception. This parameter encapsulates degradations in learning that might arise from dysfunctional synaptic plasticity mechanisms in individuals with autism. The model does not explain why some individuals perform better than others, but it does provide a simple way of modeling the outcomes of such variation, making it possible to predict how particular individuals will perform after different amounts and kinds of training. Future studies that measure the consistency of category-learning deficits within individuals across ages and across different tasks can help to identify the factors that determine which individuals show an A Type II profile.

Insights from neurally based computational models of autism

Several different neural network architectures have been developed to simulate the behavioral and perceptual deficits associated with autism (Björne & Balkenius, 2005; Cohen, 1994, 1998; Grossberg & Seidman, 2006; Gustafsson, 1997; Gustafsson & Paplinski, 2002; Noriega, 2008; O'Laughlin & Thagard, 2000; Thomas, Knowland, & Karmiloff-Smith, 2011). Of these, Grossberg and Seidman's iSTART model is the only one that directly predicts deficits in the learning of prototypes by individuals with autism. Their modeling approach is similar to ours, in that they attempted to account for individual variations in performance in terms of a systematic shift in the basic neural mechanisms underlying category learning. The iSTART model does not unambiguously predict how individuals with autism should perform in specific category-learning tasks, however. For example, Vladusich et al. (2010) used the iSTART model to qualitatively account for both the impaired and unimpaired performance in their experiments. The iSTART model assumes that dysfunctional cortical–hippocampal interactions are particularly relevant to understanding ASD-related deficits in category learning and generalization. This framework is compatible with the present simulations, if it is assumed that other brain regions, such as the hippocampus, modulate how and when neural circuits within the visual cortex change during learning.

The connectionist structure of Henderson and McClelland's (2011) ANN model makes it compatible with several more specific proposals about the involvement of cortical and subcortical brain regions in category learning (Ashby & Maddox, 2005, 2010; Nosofsky et al., 2012; Reber et al., 1998). Its success at characterizing generalization by individuals with autism suggests that focusing on

differences in cortical changes during learning may be particularly informative with respect to understanding how neural abnormalities give rise to learning deficits. For example, recent neuroimaging work has suggested that even when adults with ASD perform similarly to typical individuals, the changes in cortical activation during acquisition can differ substantially (Schipul, 2012; Schipul, Williams, Keller, Minshew, & Just, 2012). When learning to categorize dot patterns, adults with ASD showed less change in activation over time (Schipul, 2012). Specifically, when adults with ASD learned categories, activation in occipital and parietal regions remained stable or increased during acquisition, whereas in typical adults activation of these regions decreased; greater disruption of learning-induced changes in cortical activation was observed in individuals with more severe ASD symptoms. These findings strongly support the idea that abnormalities in basic neural plasticity mechanisms impact how individuals with ASD learn about perceptual categories.

The fact that a subgroup of children and adults with HFASD appears to easily learn visual categories raises questions about how atypical perceptual category learning could reflect a general neural deficit that is present in all individuals with HFASD. An alternative possibility is that heterogeneous neural abnormalities occur across individuals with HFASD that differentially affect learning across tasks. Individuals with ASD show idiosyncratic patterns of hyper- and hyporesponsiveness to inputs across modalities (Hirstein et al., 2001), consistent with this possibility. Such neural heterogeneity, if present, would be an important consideration for future structural and functional neuroimaging studies of individuals with HFASD, because if imaging data from heterogeneous subgroups are being pooled before being compared with the imaging data from typical individuals, this could provide a distorted and potentially misleading view of how neural processing differs in individuals with HFASD. Future studies that compare neural activity in autistic individuals showing an A Type I profile with activity from those showing an A Type II profile would be needed to better understand the origins of atypical perceptual generalization.

Predictions and future directions

The present simulations predict that differences in generalization between individuals with and without ASD should be most pronounced when the task requires classifying highly distorted, novel prototypes (i.e., when the task is most difficult). Consequently, the capacity of a visual category-learning task to reveal differences between individuals with and without autism may depend strongly on the range of distortions used to train and test the participants. More generally, our simulations suggest that task difficulty

may differentially degrade performance by individuals with HFASD. For instance, ANNs that generalized like A Type II individuals had a learning rate much lower than the TD model. For classification tasks in which the stimuli are easily distinguished, even large differences in learning rates will minimally impact acquisition. When the stimuli have overlapping features, however, the negative impacts of slower learning rates will become more evident. The present simulations predict that when perceptual category-learning tasks are calibrated in difficulty, such that all children learn the tasks at similar rates, differences in generalization patterns should be minimal (as reported by Vladusich et al., 2010).

Our simulations also make the surprising prediction that training participants with canonical prototypes alone (i.e., with no prototype distortions presented during training) and random stimuli should not only increase the capacity of category-learning tasks to reveal atypical perceptual categorization in children with HFASD, but may also help to identify which neural deficits are disrupting learning and generalization. Furthermore, the reduced-LR model predicts that such training may produce generalization patterns in children with HFASD that are closer to those observed in typical children who have been trained with more varied stimuli. In other words, if synaptic plasticity deficits contribute to atypical category learning, then to get children with HFASD (in particular, those with A Type II profiles) to perceptually categorize like TD children, it may be beneficial to train them with a reduced set of stimuli that do not correspond to the actual exemplars that they need to learn to classify. The idea that training with a few artificial examples could facilitate real-world generalization may seem counterintuitive, given past reports of hyperspecific learning by individuals with ASD. A potentially comparable situation is seen in infant speech learning, however, in that exposure to "motherese" can increase an infant's ability to distinguish more typical speech sounds (Liu, Kuhl, & Tsao, 2003; Werker et al., 2007). Hyperspecific learning could potentially be less problematic, and perhaps even beneficial, when the stimuli about which participants are learning are supernormal stimuli that encapsulate and/or exaggerate prototypical features of the natural stimuli.

Experimental tests of these predictions can facilitate further analysis and development of more sophisticated computational models of autism. Of the computational approaches explored here, methods instantiating dysfunctional neural plasticity or synaptic homeostasis mechanisms produced generalization patterns that were the most similar to those of children with HFASD. It seems unlikely, however, that only a subset of individuals with autism would show the effects of such neural abnormalities. Future models of category learning by individuals with autism should account for observed heterogeneities in learning

and generalization more simply and precisely. Ultimately, identifying how variations in neural substrates contribute to atypical categorical processing by individuals with ASD can facilitate the development of treatment strategies by clarifying which underlying factors may contribute to both perceptual and social deficits (Church et al., 2010; Happé & Frith, 2006; Plaisted, 2001).

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington, DC: American Psychiatric Association.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences, 5,* 204–210.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224,* 147–161.

Auerbach, B. D., Osterweil, E. K., & Bear, M. F. (2011). Mutations causing syndromic autism define an axis of synaptic pathophysiology. *Nature, 480,* 63–68.

Baruth, J. M., Casanova, M. F., Sears, L., & Sokhadze, E. (2010). Early-stage visual processing abnormalities in high-functioning autism spectrum disorder (ASD). *Translational Neuroscience, 1,* 177–187.

Behrmann, M., Thomas, C., & Humphreys, K. (2006). Seeing it differently: Visual processing in autism. *Trends in Cognitive Sciences, 10,* 258–264. doi:10.1016/j.tics.2006.05.001

Bishop, C. M. (1995). Regularization and complexity control in feedforward networks. In F. Fougelman-Soulie & P. Gallinari (Eds.), *Proceedings of the International Conference on Artificial Neural Networks (ICANN '95)* (Vol. 1, pp. 141–148). New York, NY: ACM.

Björne, P., & Balkenius, C. (2005). A model of attentional impairments in autism: First steps toward a computational theory. *Cognitive Systems Research, 6,* 193–204.

Bott, L., Brock, J., Brockdorff, N., Boucher, J., & Lamberts, K. (2006). Perceptual similarity in autism. *Quarterly Journal of Experimental Psychology, 59,* 1237–1254.

Bourgeron, T. (2009). A synaptic trek to autism. *Current Opinion in Neurobiology, 19,* 231–234.

Caruana, R., Lawrence, S., & Giles, C. (2001). Overfitting in neural networks: Backpropagation, conjugate gradient, and early stopping. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 13, pp. 402–408). Cambridge, MA: MIT Press.

Casanova, M. F., Buxhoeveden, D. P., Switala, A. E., & Roy, E. (2002). Minicolumnar pathology in autism. *Neurology, 58,* 428–432.

Casanova, M. F., van Kooten, I. A., Switala, A. E., van Engeland, H., Heinsen, H., Steinbusch, H. W., & Schmitz, C. (2006). Minicolumnar abnormalities in autism. *Acta Neuropathologica, 112,* 287–303.

Church, B. A., Krauss, M. S., Lopata, C., Toomey, J. A., Thomeer, M. L., Coutinho, M. V., & Mercado, E., III. (2010). Atypical categorization in children with high-functioning autism spectrum disorder. *Psychonomic Bulletin & Review, 17,* 862–868. doi:10.3758/PBR.17.6.862

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

Cohen, I. L. (1998). Neural network analysis of learning in autism. In D. Stein & J. Ludick (Eds.), *Neural networks and psychopathology* (pp. 274–315). Cambridge, U.K.: Cambridge University Press.

Dakin, S., & Frith, U. (2005). Vagaries of visual perception in autism. *Neuron, 48,* 497–507.

Daniel, R., Wagner, G., Koch, K., Reichenbach, J. R., Sauer, H., & Schlosser, R. G. (2011). Assessing the neural basis of uncertainty in perceptual category learning through varying levels of distortion. *Journal of Cognitive Neuroscience, 23,* 1781–1793.

Dawson, M. R. W. (2005). *Connectionism: A hands-on approach.* Oxford, U.K.: Blackwell.

Dawson, M., Mottron, L., & Gernsbacher, M. A. (2008). Learning in autism. In J. Byrne & H. L. Roediger III (Eds.), *Learning and memory: A comprehensive reference* (Cognitive psychology, Vol. 2, pp. 759–772). Oxford, U.K.: Elsevier.

Demsar, J., Zupan, B., Leban, G., & Curk, T. (2004). *Orange: From experimental machine learning to interactive data mining* [White paper]. Retrieved from www.ailab.si/orange

Fellous, J. M., & Linster, C. (1998). Computational models of neuromodulation. *Neural Computation, 10,* 771–805.

Frith, U. (1989). *Autism: Explaining the enigma.* London, U.K.: Blackwell.

Froehlich, A. L., Anderson, J. S., Bigler, E. D., Miller, J. S., Lange, N. T., DuBray, M. B., & Lainhart, J. E. (2012). Intact prototype formation but impaired generalization in autism. *Research in Autism Spectrum Disorders, 6,* 921–930. doi:10.1016/j.rasd.2011.12.006

Gastgeb, H. Z., Dundas, E. M., Minshew, N. J., & Strauss, M. S. (2012). Category formation in autism: Can individuals with autism form categories and prototypes of dot patterns? *Journal of Autism and Developmental Disorders, 42,* 1694–1704.

Gastgeb, H. Z., Rump, K. M., Best, C. A., Minshew, N. J., & Strauss, M. S. (2009). Prototype formation in autism: Can individuals with autism abstract facial prototypes? *Autism Research, 2,* 279–284. doi:10.1002/aur.93

Gastgeb, H. Z., Wilkinson, D. A., Minshew, N. J., & Strauss, M. S. (2011). Can individuals with autism abstract prototypes of natural faces? *Journal of Autism and Developmental Disorders, 41,* 1609–1618.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25. doi:10.1016/0166-2236(92)90344-8

Grossberg, S., & Seidman, D. (2006). Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates. *Psychological Review, 113,* 483–525. doi:10.1037/0033-295X.113.3.483

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

Gustafsson, L., & Paplinski, A. P. (2002). *An experiment in modelling learning in autism using self-organizing artificial neural networks (Technical Report 2001/93).* Melbourne, Australia: Monash University, School of Computer Science and Software Engineering.

Hamamoto, Y., Mitani, Y., Ishihara, H., Hase, T., & Tomita, S. (1996). Evaluation of an anti-regularization technique in neural networks. *Proceedings of the 13th International Conference on Pattern Recognition, 4,* 205–209. doi:10.1109/ICPR.1996.547416

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

Henderson, C. M., & McClelland, J. L. (2011). A PDP model of the simultaneous perception of multiple objects. *Connection Science, 23,* 161–172. doi:10.1080/09540091.2011.575931

Hirstein, W., Iversen, P., & Ramachandran, V. S. (2001). Autonomic responses of autistic children to people and objects. *Proceedings of the Royal Society B, 268,* 1883–1888.

Just, M., Cherkassky, V., Keller, T., & Minshew, N. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain, 127,* 1811–1821.

Kana, R. K., Libero, L. E., & Moore, M. S. (2011). Disrupted cortical connectivity theory as an explanatory model for autism spectrum disorders. *Physics of Life Reviews, 8,* 410–437. doi:10.1016/j.plrev.2011.10.001

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology, 13,* 111–124.

Klinger, L. G., Klinger, M. R., & Pohlig, R. L. (2007). Implicit learning impairments in autism spectrum disorders: Implications for treatment. In J. M. Perez, P. M. Gonzalez, M. L. Comi, & C. Nieto (Eds.), *New developments in autism: The future is today* (pp. 75–102). London, U.K.: Kingsley Press.

Kohonen, T. (2001). *Self-organizing maps.* Berlin, Germany: Springer.

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience, 11,* 224–231.

Lee, M. D., & Webb, M. R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review, 12,* 605–621. doi:10.3758/BF03196751

Li, S.-C., von Oertzen, T., & Lindenberger, U. (2006). A neurocomputational model of stochastic resonance and aging. *Neurocomputing, 69,* 1553–1560.

Little, D. M., & Thulborn, K. R. (2006). Prototype-distortion category learning: A two-phase learning process across a distributed network. *Brain and Cognition, 60,* 233–243.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6,* F1–F10.

Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: A review of research. *Psychological Bulletin, 86,* 1236–1254. doi:10.1037/0033-2909.86.6.1236

Markram, K., & Markram, H. (2010). The intense world theory—A unifying theory of the neurobiology of autism. *Frontiers in Human Neuroscience, 4,* 224. doi:10.3389/fnhum.2010.00224

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

Mercado, E., III. (2008). Neural and cognitive plasticity: From maps to minds. *Psychological Bulletin, 134,* 109–137. doi:10.1037/0033-2909.134.1.109

Mercado, E., III. (2011). Mapping individual variations in learning capacity. *International Journal of Comparative Psychology, 24,* 4–35.

Molesworth, C. J., Bowler, D. M., & Hampton, J. A. (2005). The prototype effect in recognition memory: Intact in autism? *Journal of Child Psychology and Psychiatry, 46,* 661–672.

Molesworth, C. J., Bowler, D. M., & Hampton, J. A. (2008). When prototypes are not best: Judgments made by children with autism. *Journal of Autism and Developmental Disorders, 38,* 1721–1730.

Moretti, P., Levenson, J. M., Battaglia, F., Atkinson, R., Teague, R., Antalffy, B., & Zoghbi, H. Y. (2006). Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *Journal of Neuroscience, 26,* 319–327.

Mottron, L., & Burack, J. (2006). Autism: A different perception. *Journal of Autism and Developmental Disorders, 36,* 1–3.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36,* 27–43. doi:10.1007/s10803-005-0040-7

Murata, N., Yoshizawa, S., & Amari, S. (1994). Network information criterion determining the number of hidden units for artificial neural network models. *IEEE Transactions on Neural Networks, 5,* 865–872.

Noriega, G. (2008). Modeling propagation delays in the development of SOMs—A parallel with abnormal brain growth in autism. *Neural Networks, 21,* 130–139. doi:10.1016/j.neunet.2007.12.006

Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences, 109,* 333–338. doi:10.1073/pnas.1111304109

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79. doi:10.1037/0033-295X.101.1.53

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255. doi:10.1111/1467-9280.00051

O'Laughlin, C., & Thagard, P. (2000). Autism and coherence: A computational model. *Mind & Language, 15,* 375–392.

Penn, H. E. (2006). Neurobiological correlates of autism: A review of recent research. *Child Neuropsychology, 12,* 57–79.

Plaisted, K. C. (2001). Reduced generalization in autism: An alternative to weak central coherence. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theories and research* (pp. 149–169). Mahwah, NJ: Erlbaum.

Plaisted, K. C., O'Riordan, M. A., & Baron-Cohen, S. (1998). Enhanced discrimination of novel highly similar stimuli by adults with autism during a perceptual learning task. *Journal of Child Psychology and Psychiatry, 39,* 765–775.

Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology, 73,* 28–38.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363. doi:10.1037/h0025953

Pothos, E. M., & Wills, A. J. (Eds.). (2011). *Formal approaches in categorization.* Cambridge, U.K.: Cambridge University Press.

Ramocki, M. B., & Zoghbi, H. Y. (2008). Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature, 455,* 912–918. doi:10.1038/nature07457

Raudys, S. (1998). Evolution and generalization of a single neurone: I. Single-layer perceptron as seven statistical classifiers. *Neural Networks, 11,* 283–296.

Raudys, S. (2001). *Statistical and neural classifiers* (1st ed.). Berlin, Germany: Springer.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory, 5,* 420–428.

Rubenstein, J. L. R., & Merzenich, M. M. (2003). Model of autism: Increased ratio of excitation/inhibition in key neural systems. *Genes, Brain, and Behavior, 2,* 255–267.

Samson, F., Mottron, L., Jernel, B., Belin, P., & Ciocca, V. (2006). Can spectro-temporal complexity explain the autistic pattern of performance on auditory tasks? *Journal of Autism and Developmental Disorders, 36,* 65–76.

Samson, F., Mottron, L., Soulières, I., & Zeffiro, T. A. (2012). Enhanced visual functioning in autism: An ALE meta-analysis. *Human Brain Mapping, 33,* 1553–1581. doi:10.1002/hbm.21307

Schipul, S. E. (2012). *Neural mechanisms of implicit and explicit learning in autism.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Schipul, S. E., Williams, D. L., Keller, T. A., Minshew, N. J., & Just, M. A. (2012). Distinctive neural processes during learning in autism. *Cerebral Cortex, 22,* 937–950.

Schmeisser, M. J., Ey, E., Wegener, S., Bockmann, J., Stempel, A. V., Kuebler, A., & Boeckers, T. M. (2012). Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature, 486,* 256–260.

Schreibman, L., Koegel, R. L., & Craig, M. S. (1977). Reducing stimulus overselectivity in autistic children. *Journal of Abnormal Child Psychology, 5,* 425–436.

Schroeder, J. H., Desrocher, M., Bebko, J. M., & Cappadocia, M. C. (2010). The neurobiology of autism: Theoretical applications. *Research in Autism Spectrum Disorders, 4,* 555–564.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science, 249,* 892–895. doi:10.1126/science.2392679

Sheppard, E., Ropar, D., & Mitchell, P. (2009). Drawing the line: How people with autism copy line drawings of three-dimensional objects. *Perception, 38,* 1104–1106.

Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research, 49,* 2705–2739.

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 984–1002. doi:10.1037/0278-7393.27.4.984

Smith, J. D., Murray, M. J., Jr., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 659–680. doi:10.1037/0278-7393.23.3.659

Solomon, M., Smith, A. C., Frank, M. J., Ly, S., & Carter, C. S. (2011). Probabilistic reinforcement learning in adults with autism spectrum disorders. *Autism Research, 4,* 109–120.

Soulières, I., Mottron, L., Giguère, G., & Larochelle, S. (2011). Category induction in autism: Slower, perhaps different, but certainly possible. *Quarterly Journal of Experimental Psychology, 64,* 311–327. doi:10.1080/17470218.2010.492994

Soulières, I., Mottron, L., Saumier, D., & Larochelle, S. (2007). Atypical categorical perception in autism: Autonomy of discrimination? *Journal of Autism and Developmental Disorders, 37,* 481–490. doi:10.1007/s10803-006-0172-4

Spencer, J., O'Brien, J., Riggs, K., Braddick, O., Atkinson, J., & Wattam-Bell, J. (2000). Motion processing in autism: Evidence for a dorsal stream deficiency. *NeuroReport, 11,* 2765–2767.

Thomas, M. S. C., Knowland, V. C. P., & Karmiloff-Smith, A. (2011). Mechanisms of developmental regression in autism and the broader phenotype: A neural network modeling approach. *Psychological Review, 118,* 637–654.

Vandenbroucke, M. W. G., Scholte, H. S., van Engeland, H., Lamme, V. A. F., & Kemner, C. (2008). A neural substrate for atypical low-level visual processing in autism spectrum disorder. *Brain, 131,* 1013–1024.

Vattikuti, S., & Chow, C. C. (2010). A computational model for cerebral cortical dysfunction in autism spectrum disorders. *Biological Psychiatry, 67,* 672–678.

Vladusich, T., Olu-Lafe, O., Kim, D. S., Tager-Flusberg, H., & Grossberg, S. (2010). Prototypical category learning in high-functioning autism. *Autism Research, 3,* 226–236.

Wang, G., Grone, B., Colas, D., Appelbaum, L., & Mourrain, P. (2011). Synaptic plasticity in sleep: Learning, homeostasis and disease. *Trends in Neurosciences, 34,* 452–463.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition, 103,* 147–162.

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin, 138,* 102–125. doi:10.1037/a0025715

Yechiam, E., Arshavsky, O., Shamay-Tsoory, S. G., Yaniv, S., & Aharon, J. (2010). Adapted to explore: Reinforcement learning in autistic spectrum conditions. *Brain and Cognition, 72,* 317–324.

Yizhar, O., Fenno, L. E., Prigge, M., Schneider, F., Davidson, T. F., O'Shea, D. J., & Deisseroth, K. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature, 477,* 171–178.