A learning-style theory for understanding autistic behaviors

Ning Qian[1,2]* and Richard M. Lipkin[1]

[1] Department of Neuroscience, Columbia University, New York, NY, USA
[2] Department of Physiology and Cellular Biophysics, Columbia University, New York, NY, USA

Understanding autism's ever-expanding array of behaviors, from sensation to cognition, is a major challenge. We posit that autistic and typically developing brains implement different algorithms that are better suited to learn, represent, and process different tasks; consequently, they develop different interests and behaviors. Computationally, a continuum of algorithms exists, from lookup table (LUT) learning, which aims to store experiences precisely, to interpolation (INT) learning, which focuses on extracting underlying statistical structure (regularities) from experiences. We hypothesize that autistic and typical brains, respectively, are biased toward LUT and INT learning, in low- and high-dimensional feature spaces, possibly because of their narrow and broad tuning functions. The LUT style is good at learning relationships that are local, precise, rigid, and contain little regularity for generalization (e.g., the name–number association in a phonebook). However, it is poor at learning relationships that are context dependent, noisy, flexible, and do contain regularities for generalization (e.g., associations between gaze direction and intention, language and meaning, sensory input and interpretation, motor-control signal and movement, and social situation and proper response). The LUT style poorly compresses information, resulting in inefficiency, sensory overload (overwhelm), restricted interests, and resistance to change. It also leads to poor prediction and anticipation, frequent surprises and over-reaction (hyper-sensitivity), impaired attentional selection and switching, concreteness, strong local focus, weak adaptation, and superior and inferior performances on simple and complex tasks. The spectrum nature of autism can be explained by different degrees of LUT learning among different individuals, and in different systems of the same individual. Our theory suggests that therapy should focus on training autistic LUT algorithm to learn regularities.

Keywords: autism spectrum disorder, Asperger's, synaptic plasticity, over selectivity, anxiety, fear, redundancy, modeling

INTRODUCTION

Autism is defined behaviorally. Since Kanner's (1943) classic paper, both psychophysical experiments and case reports have documented a diverse array of autistic behaviors, ranging from sensory perception and motor control, to learning, memory, language, and social interaction, and from inferior to superior performances. The following is a partial list, starting with the triad used in diagnosis (DSM-IV-TR).

1. Impaired social interaction. Apparent more interest in objects than people.
2. Impaired communication skills. Rigid and limited use of language.
3. Stereotyped behaviors. Repetitive movements, restricted interests, and insistence on sameness (rituals).
4. Preference of non-social stimuli (such as audio–visual synchrony of clapping hands) to social ones (such as eyes, faces, biological motion; Osterling and Dawson, 1994; Klin et al., 2009).
5. Easily overwhelmed in social situations or crowded places (Kanner, 1943; Grandin, 2006; Robison, 2011).
6. Superior lexical processing of individual words in a picture naming task (Walenski et al., 2008) but difficulty with grammar (Waleski et al., 2006).

7. Poor predictive ability. Failure to anticipate picking-up by parents (Kanner, 1943), timing of air puffs to eyes (Sears et al., 1994), and future position of moving objects (Sinha, 2011).
8. Hyper- and hypo-sensitivity to sensory stimuli (Kanner, 1943; Volkmar et al., 1986; Oneill and Jones, 1997).
9. Atypical learning. Resistance to conventional training, but spontaneous learning of atypical tasks, such as memorizing a phonebook (Dawson et al., 2008).
10. Different patterns of generalization in perceptual learning (Plaisted et al., 1998), motor learning (Haswell et al., 2009), and word-list memory tasks (Beversdorf et al., 2000).
11. Preference of concreteness to abstraction (Kanner, 1943; Grandin, 2006).
12. Difficulty in attentional selection or switching (Courchesne et al., 1994; Grandin, 2006). Hyper-focus on tasks of interest (Robison, 2011).
13. Focus on local features instead of global patterns. Seeing "trees" but not "forest." Superior performance on local tasks. Reduced contextual modulation or interference (Frith and Happe, 1994; Baron-Cohen, 2002; Happe and Frith, 2006).
14. Impaired recognition of facial emotions and face identity (Weeks and Hobson, 1987; Boucher and Lewis, 1992).

15. Reduced face-identity aftereffects after face adaptation (Pellicano et al., 2007).
16. Impaired motion processing, particularly second-order motion and biological motion (Kaiser and Shiffrar, 2009).
17. Better and worse performances on simple and complex tasks, respectively, compared with controls (Minshew and Goldstein, 1998; Bertone et al., 2005; Mottron et al., 2006). Faster learning of simple eye-blink conditioning (Sears et al., 1994).
18. Superior absolute pitch judgment (Heaton et al., 1998).

Dozens of autism theories, at levels from anatomy to cognition, have already been proposed. Although each has its strengths, most theories focus on only a subset of the behaviors and cannot explain the majority of them. The lack of a unified account greatly impedes efforts to define autism and link its behaviors to underlying physiology, anatomy, and ultimately, genetics. Fragmented accounts of autistic behaviors sometimes lead to opposite therapeutic recommendations.

We propose a theory to explain a wide variety of autistic behaviors. Based on the well-known difficulty of training autistic people (Dawson et al., 2008), we suggest that autistic and typical brains are biased toward different learning styles that are better suited to learn different tasks, and consequently, follow different developmental paths. Since learning is intimately linked to development, processing, and representation, and occurs in all brain regions, at all processing levels, and across all stages of life, this learning-style theory matches autism's pervasive nature. The theory also provides a potential link to underlying physiological mechanisms, such as altered neuronal selectivity and synaptic plasticity.
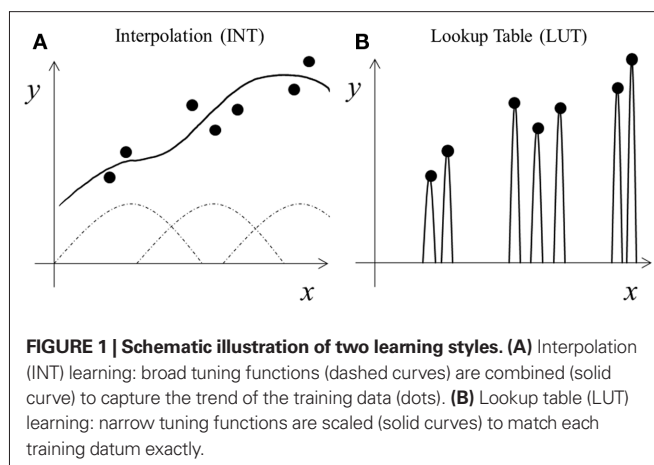
THEORY

Like previous autism theories, our theory is qualitative. Since autistic behaviors range from sensory perception and motor control to language and social interaction, quantitative simulations would require a plausible neural model for each affected system. While future research should aim at this formidable goal, we show in this paper that qualitative reasoning in a learning framework logically explains most autistic behaviors.

There are many computational learning theories and algorithms. Our framework does not depend on a specific theory; rather, it depends on a tradeoff between two learning styles. For concreteness, we consider a generic supervised learning task of mapping input x to output y (**Figure 1**). Both x and y can be multi-dimensional although they are only one dimensional in **Figure 1** for easy illustration. For instance, x may represent the neuronal control signals for various arm muscles and y the resulting hand displacement vector. The goal is to learn the input–output mapping from some training examples (dots in **Figure 1**). Each example means that a specific x is known to produce a specific y according to experience.

LOOKUP TABLE (LUT) VS. INTERPOLATION (INT) STYLES OF LEARNING
Among other things, learning concerns how well to represent training examples (dots in **Figure 1**) and how to handle new cases (gaps between the dots). Specifically, learning is about fitting training data with "proper" functions that generalize "best" to new data drawn from the same distribution as the training examples (Poggio and Girosi, 1990). Most machine learning theories analyze the tradeoff

FIGURE 1 | Schematic illustration of two learning styles. (A) Interpolation (INT) learning: broad tuning functions (dashed curves) are combined (solid curve) to capture the trend of the training data (dots). **(B)** Lookup table (LUT) learning: narrow tuning functions are scaled (solid curves) to match each training datum exactly.

between the complexity of the fitting functions and the generalization performance: more complex functions fit training data better but require more training examples to generalize properly (Poggio and Girosi, 1990; Shawe-Taylor and Cristianini, 2000). We consider a somewhat different tradeoff because biological systems do not have complete freedom to choose any fitting functions; instead, they rely on neurons' response properties typically characterized by their tuning along various dimensions (space, time, orientation, motion, disparity, color, shape, face configuration, control signals, etc.). Many cells are jointly tuned to multiple dimensions, and their tuning is characterized by high-dimensional surfaces. Learning, then, involves combining these tuning surfaces to fit training examples and interpolate across gaps between training examples for generalization (curve in **Figure 1**; Poggio and Girosi, 1990). This maybe done in multiple stages, as in multi-layer neural networks. We interpret tuning most generally to include the range of stimuli that influence a cell's response in any way (see Discussion).

We now contrast the two learning styles for our autism theory. The interpolation (INT) style uses broad tuning functions to fit the training examples and interpolate across gaps between examples (**Figure 1A**). It aims not to represent each training example precisely, but to find underlying trends or statistical regularities from training data in order to generalize. If training data cluster, then using just one tuning function for each cluster is sufficient (Poggio and Girosi, 1990), enabling regularity extraction with minimum neuronal resources. In contrast, the lookup table (LUT) style of learning aims to store each training example precisely, but ignores underlying trends for generalization. One way to do so is to assign a narrow tuning function to each training example (**Figure 1B**). The narrow tuning reduces interference between nearby inputs, making it easier to represent individual training examples precisely, but does not generalize well to gaps between training examples. Varying the tuning widths produces a continuum of learning algorithms from INT to LUT.

Note that even with broad tuning for INT learning, a multi-layer network with sufficient units can approximate any continuous input–output mapping arbitrarily well (Funahashi, 1989; Girosi and Poggio, 1990), and thus can fit any finite training data arbitrarily well. However, the INT style emphasizes extracting underlying statistical regularities from training data without aiming to fit the data precisely. First, if training data are noisy (which is often the case), then fitting them precisely is a pointless waste of neuronal resources.

Second, after fitting a set of training examples precisely, adding new examples will compromise that precision because of interferences between the new and old examples through broad tuning functions, so that all the examples must be re-trained together. Third, if a task permits small errors or can use sensory feedback to correct them, then it is not necessary to learn the mapping precisely at a great cost.

In **Figure 1**, we consider learning a continuous input–output mapping. The same discussion applies to classification or categorization tasks, in which INT learning reduces to finding continuous decision boundaries to divide the input space into separate categories (**Figure 2**). The discussion also extends to unsupervised learning where input data do not have explicit output labels. Here, INT learning clusters or compresses input data based on their statistical regularities. In both cases, LUT learning aims to store individual training examples precisely without much processing.

Our main hypothesis is that typical and autistic brains are biased toward INT and LUT styles of learning, respectively.

OVER-FITTING CANNOT EXPLAIN AUTISM
We explained above a LUT learning style with narrow tuning functions. An alternative is to use functions of larger-than-necessary capacity or complexity (Shawe-Taylor and Cristianini, 2000) to learn a given task. This results in precise, over-fitting of training examples (**Figure 3**), and may also be regarded as LUT learning. A related idea is Cohen's (1994) assumption that autistic brains

have a larger-than-necessary number of units/connections. The large difference between the narrow tuning and over-fitting versions of LUT learning can be seen by comparing **Figures 1B and 3**. The former hardly generalizes to gaps between training examples and shows little interference among training examples whereas the opposites are true for the latter.

We feel that the over-fitting version of LUT learning (**Figure 3**) does not match autistic behaviors well because the impact of over-fitting on generalization performance can be eliminated by using more training examples (the number grows only linearly with a capacity measure called VC dimension; Shawe-Taylor and Cristianini, 2000). Indeed, Cohen's (1994) simulations suggest that the difference in generalization performances between his networks with and without over-fitting is too small to account for the devastating nature of severe autism (also note the small difference between **Figures 1A and 3**). Additionally, the over-fitting scheme has interference between new and old training examples, and may not be consistent with autism's strong memorization of random facts (Dawson et al., 2008). Finally, if autistic people used functions of larger capacity, then with sufficient training examples they would eventually learn any complex task better than controls, which does not appear to be true.

In the following, the term LUT learning always refers to the narrow tuning version unless otherwise specified.

TUNING WIDTH, FEATURE-SPACE DIMENSIONALITY, AND TASK COMPLEXITY
For many real-world problems, input–output mappings are convoluted. For example, the raw visual inputs are luminance values on a two-dimensional image, and the output may be face recognition of the image. Since large input variations (e.g., different views of the same person or different lighting conditions) may correspond to the same output and small input variations (e.g., same front view of different persons) may correspond to different outputs, the mapping cannot be fitted by a simple function. A solution is first to project the raw input space into an appropriate input feature space where even a linear function can solve the problem (Poggio and Girosi, 1990; Shawe-Taylor and Cristianini, 2000). Indeed, intermediate layers of multi-layer neural networks represent feature spaces. Biological systems appear to work similarly: neurons along the visual hierarchy are tuned to various features (space, time, orientation, motion, disparity, color, shape, face configuration, etc.) instead of raw luminance values. Thus, in **Figures 1–3**, the x axes really represent relevant input features instead of raw inputs, and the relevant feature set depends on the task.

This raises the possibility that autistic and typical brains might learn in different feature spaces or with different efficiency of feature representations (e.g., explicit features vs. their inner products; Shawe-Taylor and Cristianini, 2000). While future work should examine this issue in detail, here we focus on an implication of the tuning-width assumption above, namely that autistic LUT learning and typical INT learning favor low- and high-dimensional feature spaces, respectively. Specifically, Fisher information, which bounds accuracy of unbiased estimators, scales with the tuning width (w) according to:

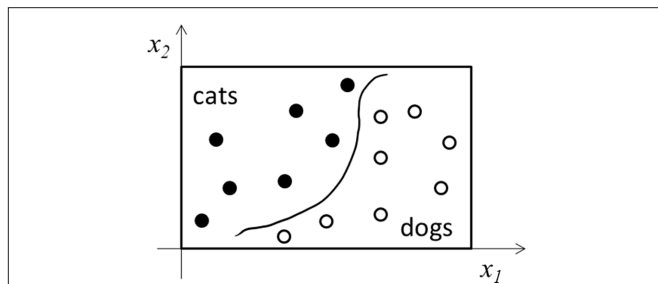$$I_F \propto w^{d-2},\hspace{4cm}(1)$$

FIGURE 2 | Schematic illustration of a categorization task. Only two input dimensions ($x1$ and $x2$) and two categories (cats and dogs) are shown. INT learning finds the decision boundary (curve) as a regularity of the training examples (filled and open dots). LUT learning stores precisely the examples without extracting regularity.
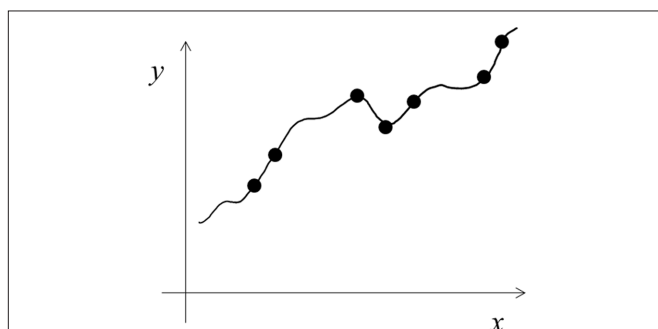

FIGURE 3 | Schematic illustration of over-fitting using functions of larger-than-necessary capacity. This version of LUT learning cannot explain autism (see text) and will not be considered further in this paper.

where d is the dimensionality of the feature space (Zhang and Sejnowski, 1999). (Fisher information of a single cell scales with w^{-2} but the number of contributing cells scales with w^d, Abbott and Dayan, 1999.) This result suggests that to maximize Fisher information for learning, representation, and processing, narrow and broad tuning functions of the LUT and INT styles favor low- and high-dimensional feature spaces, respectively. Intuitively, when tuning is narrow, using a higher-dimensional feature space does not really increase the number of contributing cells, and the system is better off coding each dimension separately than coding all dimensions combinatorially.

An equivalent conclusion is that LUT and INT styles are better at learning simple and complex tasks, respectively. For simple tasks involving a single feature (e.g., first-order orientation discrimination), $d = 1$ in Eq. 1, and Fisher information is greater for narrower tuning (Zhang and Sejnowski, 1999) of autistic LUT style. For second-order orientation discrimination, however, the brain has to represent properties (such as texture and contrast modulation) that define the orientation, in addition to orientation itself. For even more complex tasks such as face recognition, a much higher-dimensional feature space is required (e.g., sizes and shapes of eyes, nose, mouth, contours, and spatial relationships among them) so that $d \gg 1$, and Fisher information is greater for broader tuning (Zhang and Sejnowski, 1999) of typical INT style. (Note, however, that very broad tuning is not desirable based on energetic considerations; Zhang and Sejnowski, 1999.)

EXPLANATION OF AUTISTIC BEHAVIORS

We hypothesize that autistic and typical brains are biased toward LUT and INT learning, respectively. The LUT and INT styles may be realized, respectively, by narrow and broad neuronal tuning functions, and by a strong and weak emphasis on coding training data precisely. We now show that by examining the logical consequences of this hypothesis, we can explain the broad array of autistic behaviors listed in the Introduction. For the ease of discussion we often consider the extreme case of LUT learning. However, we do *not* imply that every autistic individual has the same set of behavioral characteristics of the same severity (see Spectrum Nature of Autism).

IMPLICATIONS OF THE LUT VS. INT STYLES OF LEARNING

The LUT style is well suited to learn tasks that are local, precise, rigid, and contain little inherent structure or regularity for generalization. Its local preference arises for two related reasons. First, LUT learning relies on narrower tuning functions and thus covers a smaller scale in each affected dimension. Second, narrower tuning favors using a lower-dimensional feature space (see Theory); this makes LUT learning more context-independent (thus more local) because contexts are included by adding extra dimensions to the feature space. For example, the saliency of a contour segment depends on how it aligns with nearby segments (Werthermer, 1938; Li, 2000). This context dependence can be represented by adding feature dimensions to represent the nearby segments. The LUT style also prefers precise, rigid relationships because it aims to store training data precisely. Finally, since LUT learning does not generalize, it well matches tasks that contain little inherent

regularity for generalization. A perfect task for autistic LUT learning is to memorize a phonebook. The name–number association in a phonebook is precise and context independent, and has little regularity for generalization: one cannot interpolate persons A and C' numbers to determine person B's number. Moreover, narrow tuning minimizes interferences between new and old numbers.

In contrast, the LUT style is poor at learning tasks that are context dependent, noisy, flexible, and do contain inherent structure for generalization; these tasks are better handled by the INT learning style. The broader tuning functions of the INT style cover larger scales and encourage using higher-dimensional feature spaces, leading to more global and context-dependent processing. The INT style focuses on extracting underlying regularities from training data, and is thus good at learning tasks with regularities. Once the regularities are learned, the system can use them to generalize to new cases and gain efficiency. The INT style also handles noisy training data well because it does not insist on fitting details, which could be noise.

Context dependence is a form of statistical regularity. The contour-saliency example above results from the fact that statistically, nearby contour segments belonging to the same object boundary in the real world tend to align smoothly (Geisler et al., 2001; Sigman et al., 2001). This is another reason that the INT style, which extracts regularity, shows more context dependence than the LUT style, which does not extract regularity.

IMPAIRED SOCIAL INTERACTION

The relationships between social situations and proper responses are noisy, flexible, and context dependent, and contain underlying (albeit fuzzy) regularities for generalization. For example, general "rules" govern how people respond when they meet someone, but specifics are never precise and depend on context such as whether the new person is a casual acquaintance or a potential employer, employe, mate, etc. Likewise, when old friends gather, handshakes, hugs, pats, etc., are bound to happen, but exactly, say, how tight a hug will be, is variable and context dependent. According to our theory, these are the relationships that are difficult for the autistic LUT style to learn. The LUT style tries to store each social experience by rote. However, without extracting complex regularities in a high-dimensional feature space, LUT learning cannot effectively use the stored information to generalize to new, related situations. The best autistic people could do is to follow rigidly the memory entry that best matches the current situation as a script. Grandin, an autistic author, has written about her handling social situations better as she gets older because she accumulates more examples in her "visual library" and presumably can find a better match to each social situation (Grandin, 2006).

Compared with controls, autistic people look at others' eyes much less frequently (Osterling and Dawson, 1994; Dalton et al., 2005) and have little task-dependent modulation in the brain's gaze areas such as superior temporal sulcus (Pelphrey et al., 2005). Part of the reason, according to our theory, is that the relationship between gaze direction and intention is hard for autistic LUT style to learn. By looking at a baby, a parent, nearby or across the room, may pick her up soon or some time later, may feed her or move her to a new spot, or may simply check her safety without further action; exactly what will happen is variable and depends on the context

such as whether the baby is crying or was fed recently. LUT learning attempts to store each instance separately and precisely and fails to extract the fuzzy, context-dependent relationship between gaze direction and intention. In addition, the broad range of spatial and temporal scales involved in this relationship is hard for narrowly tuned neurons to represent. Other reasons for autistic people's gaze aversion are that face is a complex, dynamic stimulus that may overload the inefficient LUT system and that low-dimensional feature spaces of the LUT style may not represent subtle facial emotions and cues well (see below).

On the other hand, the LUT style can readily learn simple associations or correlations, such as those between a tone and an air puff to the eye (Sears et al., 1994), and between visual and auditory inputs created by clapping hands (Klin et al., 2009), because these relationships are precise, reliable, and local in space and time in the cited experiments.

Thus, our theory explains why autistic babies attend more to visuo-audial synchronies (such as clapping hands) than to socially relevant stimuli (such as eyes, faces, and biological motion). The standard interpretation is that this early preference of non-social to social stimuli places autistic people on a different developmental path, and leads to full blown autism (Klin et al., 2009). Our theory further suggests that the reason for autistic preference of non-social to social stimuli is that most socially relevant relationships happen to be hard for the LUT style to learn. If a baby cannot learn the information in the gaze, then he/she will be less interested in looking at eyes, which further reduces his chance of learning the information in the gaze.

IMPAIRED LANGUAGE AND COMMUNICATION

It is long recognized that lexical processing of individual words is relatively spared in autism (Walesski et al., 2006). Using less frequently used words in a picture naming task, Walenski et al. (2008) recently showed that autistic word processing is not only spared but also superior to controls. However, autistic people have great difficulty learning grammatical rules and context-dependent use of words (Happe and Frith, 2006; Walesski et al., 2006). Our theory explains these findings. The definitions of individual words are largely random associations between words and other words or objects in the world. In Walenski et al.'s (2008) experiment, the association was between words and pictures. Autistic LUT style can learn these associations well because they are fixed and involve little generalization. Even for words with multiple meanings, their long definitions are not context dependent; their usage in a sentence is.

In contrast, flexible and competent language production and understanding require using words in a grammatically correct and context sensitive way. The autistic LUT style is poor at extracting grammatical rules and context dependence, which are regularities hidden in the training examples (e.g., utterances from parents). Moreover, low-dimensional feature spaces of the LUT style impair context-dependent processing (see above). A recent airline advertisement, "Flights, hotels, wheels and more," illustrates the point. In this context, "wheels" means "cars." However, a system with LUT learning, which stores all the definitions of "wheels" without using additional feature dimensions to represent context, would hardly be able to pick "cars" as the unique answer. Indeed, cultural, social, and situational contexts are all important for language. Poor context

representation explains autistic people's literal (Happe and Frith, 2006) or "logical" (Robison, 2011) use of words. Because of their weak learning of grammar and context as regularities for generalization, yet strong memorization of detailed examples, autistic people rely on memorized sentences as scripts, resulting in formulaic speech.

Walenski et al. (2008) propose that words and grammar are processed by the declarative and procedural memory systems, respectively, and that autistic people have a defective procedural memory system, and hence impaired grammar. Comparing their and our explanations raises the possibility that in a given brain, autistic or not, the declarative memory system, which has to store random facts, may be more biased toward the LUT style, whereas the procedural memory may be more biased toward the INT style. However, the declarative system is not just for random facts but also for coherent stories with underlying structures. Walenski et al.'s (2008) theory predicts that autistic people tell coherent stories normally because of their intact declarative system. Our theory, in contrast, predicts that autistic people's LUT bias makes it harder for them to learn any complex regularity, including structures of coherent stories. Relevant experiments appear to support the latter prediction (Losh and Capps, 2003; Diehl et al., 2006) although further studies are needed.

IMPAIRED INFORMATION COMPRESSION, INEFFICIENCY, SENSORY OVERLOAD, AND OVERWHELM

Information processed by the brain is often highly redundant, so a common processing strategy is to compress information by reducing redundancies to improve coding efficiency (Barlow and Foldiak, 1989; Zhaoping, 2006). A prerequisite of compression, however, is to discover underlying statistical structure or regularity in the data or task. Indeed, regularity is redundancy. Real data often reside on a low-dimensional manifold of a high-dimensional raw input space, and learning underlying regularity for data compression amounts to finding this manifold by properly interpolating training data (Roweis and Saul, 2000; Tenenbaum et al., 2000). Moreover, regularity generalizes to new data (drawn from the same distribution as the training data), making their representation, and processing more efficient as well. Additionally, coordinates of the manifold define useful features for learning other tasks such as classification.

Since autistic LUT style is poor at learning regularity, it cannot thoroughly compress input information to remove redundancy and define useful features, and is thus inefficient and resource intensive. The consequent information overload explains why autistic people may easily become overwhelmed in social or public places (Kanner, 1943; Grandin, 2006; Robison, 2011) where there is a wealth of sensory stimulation. Sensory information is both rich and redundant. The raw visual inputs, for example, are luminance values sensed by millions of retinal cones. This enormous input space, however, is highly structured and redundant at multiple levels. At a low level, nearby points in space and time tend to have similar luminance values because of surfaces in the world. At a high-level, object shapes are invariant to changing lighting conditions and retinal positions. The typical visual pathway has mechanisms, from retina (e.g., center-surround receptive fields) to inferior temporal cortex (e.g., relatively position-invariant tuning), to exploit these regularities and compress information. Autistic LUT style, being poor at

learning regularities, must fail to develop some of the mechanisms fully, and likely be overwhelmed by sensory overload. Higher-level mechanisms may be more affected because cascaded processing along a hierarchy may compound and amplify the deficit of LUT learning at higher levels.

Moreover, poor compression and inefficient coding must affect complex stimuli more than simple ones because complex stimuli are information rich. For example, faces, with their many dynamic parts, contain more information than oriented bars or gratings. Motion stimuli contain more information than static ones. This may contribute to autistic people's difficulty with face and motion processing, and indeed with complex stimuli in general. A typical brain might use dozens of relevant features to compress and represent retinal images efficiently whereas an autistic brain might extract only a few simple features which cannot capture regularity and remove redundancy well, and might have to store images in relatively raw forms. Autistic people's looking away from faces and biological motion may be a strategy of reducing information overload by using lower-resolution periphery vision. Another reason is that the LUT style is poor at learning socially relevant signals in these stimuli (see Impaired Social Interaction above).

RESTRICTED INTEREST, REPETITIVE BEHAVIOR, RESISTANCE TO CHANGE, AND TALENTS

Although we focused on sensory processing above, the argument applies to other brain functions. Language, for example, is redundant at multiple levels from phonology to syntax (Darian, 1979). Learning these regularities is far more efficient than memorizing a great number of example sentences. In motor control, the relationship between muscle forces and movements are determined by Newtonian mechanics. Learning to approximate this relationship is far more efficient than to store a huge number of example movements. By spending too much resource on highly redundant information, the LUT style can learn only a limited set of behavioral repertoires, leading to restricted interests, limited language ability, repetitive behaviors, and resistance to change. Autistic people's insistence on repeating the same rituals may be partly a strategy of reducing information overload by avoiding new information. (Their social anxiety, another consequence of LUT learning (see below), must also contribute to this behavior.) The inefficiency of the LUT style could further force autistic brains to use low-dimensional feature spaces for learning and processing.

The above argument does not contradict observations that autistic people can memorize random facts, such as phone numbers, and a small fraction of them even has special talents in tasks such as calculating calendar or solving Rubik's cube (Baron-Cohen et al., 2009). First, the information content and computational complexity (and thus neuronal resource requirement) for these "modern" tasks are far dwarfed by those for relatively "primal" tasks (sensorimotor processing, language, and social interaction). Minsky (1985) summarized this counter-intuitive fact as "easy things are hard." Because primal tasks are essential for survival over the long evolutionary history, the brain has sophisticated machineries to handle them and they appear easy. In contrast, modern tasks (chess, mathematics, physics, engineering, programming, drawing as well as calendar calculation and Rubik's cube) impact survival *only* after the dawn of civilization. Consequently, evolution must not have had enough time to perfect

their processing by the brain and they appear hard. The incorrectness of common intuition is testified by the fact that off-the-shelf software solves math problems better than overwhelming majority of people while no state-of-the-art robot remotely approaches visuo-motor capabilities of an average 3-year old. Second, modern tasks are usually more precise and rigid than primal tasks, and in this aspect, well match autistic LUT learning. (However, some modern tasks such as mathematics involve complex regularities and generalization that are beyond precise-rule-based computation. Although people with autistic traits of precision and focus have advantages, those with severe autism must have difficulty with such tasks.)

OVER TRAINING ON NOISY DATA

Autistic LUT learning aims to store training data precisely. This is advantageous when the task is to memorize a phonebook; a single digit error will render a phone number useless. However, most relationships in sensorimotor processing and social interaction are noisy. For example, the same motor command does not produce exactly the same movement. The LUT style considers any fluctuations in an input–output mapping as errors to eliminate, in a constant, futile effort to chase noise. This leads to over training on a limited set of behavioral repertoires (see above), making it hard to break the learned habits. For memories stored in recurrent networks (Hopfield, 1982), over training produces strong attractor states (deep wells on energy landscapes) or state sequences, in which the system may be trapped for a long time (repetitive behaviors).

Cohen (1994) first used over training to explain autism, but in a neural network with a larger-than-necessary number of units/connections to overfit data. As noted in the Section "Theory," the over-capacity and over-fitting assumption does not seem to match autism well. In our theory, over training is a consequence of the LUT style's insistence on storing fluctuating data precisely with narrow tuning functions instead of over-fitting with a network of larger-than-necessary capacity.

ATYPICAL LEARNING

Autistic people are hard to train in typical social behaviors; yet, they spontaneously learn things that typical people consider hard, such as memorizing license-plate numbers of parked cars or phonebooks (Dawson et al., 2008). Our theory explains such observations trivially because as noted above, autistic LUT, and typical INT styles are better suited to learn different tasks.

Our theory suggests that autistic people's difficulty of learning social behavior is similar to typical people's difficulty of memorizing random factual details, such as phone numbers; both arise from a mismatch between learning style and task. Intensive, brute force training would surely make typical people remember a phonebook better but they will never do quite as well as autistic people. In particular, although broad tuning functions of the INT style are good for learning complex regularities in high-dimensional feature spaces, they cause interference between new and old numbers, and the old numbers will have to be re-trained constantly. Likewise, intensive, brute force training would help autistic people by supplying more examples to match a given situation but they will never have the flexibility and efficiency afforded by intricate regularities learned by typical people. In Section "Therapeutic Implications," we suggest some alternative training strategies.

POOR GENERALIZATION OF WORD-LIST, PERCEPTUAL, AND MOTOR LEARNING, AND SOME SUPERIOR PERFORMANCES

Our theory also trivially explains autistic people's poor generalization of learning to new situations because the LUT style does not interpolate training data to generalize (**Figure 1B**). As discussed above, this inability of regularity learning and generalization contributes to many autistic behaviors. Here we explain findings from a few learning studies, including superior performances, as reduced generalization.

Beversdorf et al. (2000) asked subjects to remember a list of words read to them, e.g., thread, pin, eye, sewing, sharp, point, prick, thimble, etc. They then read subjects another list, e.g., thread, pie, needle etc., and subjects had to report whether each word was on the first list. The target words on the second list were those that were not on, but semantically related to, the first list (e.g., needle). They found that autistic subjects had fewer false reports than controls on the target words. Our theory explains this superior performance as reduced generalization in the semantic space. Typical INT learning clusters words of related meanings in the semantic space; this form of regularity extraction and generalization facilitates common use of language but leads to false reports in this particular task. In contrast, autistic LUT learning does not generalize and thus shows less interaction among related words.

Plaisted et al. (1998) trained subjects to discriminate a pair of visual patterns (trained pair). Both autistic and control subjects improved their performances with training sessions. They then tested subjects' discriminability on two new pairs of patterns (test pairs). The first test pair was similar to the trained pair, and the second test pair was more different. For control subjects, learning the trained pair transferred positively and negatively to the first and second test pairs, respectively, whereas for autistic subjects, neither positive nor negative generalization occurred, as predicted by our theory. Plaisted et al. (1998) emphasized that autistic subjects performed better than controls for the second test pair. We explain this superior performance as reduced negative generalization.

As an aside, we note that the controls' specific pattern of transfer is explained by Teich and Qian's (2003) perceptual-learning model with typical tuning widths (see their Figure 7B). Although that model concerns orientation, the idea is generally applicable: learning changes neuronal tuning by shifting resources toward the trained pair to improve their discrimination. When the tuning functions are broad, this resource shift affects similar stimuli positively and more different stimuli negatively. However, if the tuning is very narrow, as we assume for autistic LUT style, then learning is local in the feature space with neither positive nor negative transfer to other stimuli.

Haswell et al. (2009) studied motor learning in a force field perpendicular to movement direction. Both autistic and control subjects learned to produce a lateral force to counter the force field, in the left workspace. The angle between the movement direction and the forearm orientation (movement-to-forearm angle) was 90° during training. Subjects were then tested in the right workspace. When the movement-to-forearm angle was 135°, autistic subjects produced little lateral force compared with the controls. We explain this finding as autistic subjects' poor generalization to the new configuration. However, when the movement-to-forearm angle was

90°, autistic subjects produced stronger lateral force than controls. This may result from two factors. First, since the 90° test angle was identical to that used for training, autistic LUT style better memorizes the required lateral force for this angle. Second, the original force field used in training was *not* applied during testing; instead, a force channel clamped movements on a straightline to target, regardless of subjects' lateral force. Since controls could learn the new rule (no need to produce a lateral force) better than autistic subjects, they produced weaker lateral forces.

Haswell et al. (2009) explained their findings by assuming that autistic subjects have a defective internal model that relies more on proprioception than vision. Their and our explanations are not mutually exclusive. In fact, internal models are a regularity that typical people learn to predict sensory consequences of their movements. Autistic subjects are poor at learning regularities and must have defective internal models.

CONTEXT INDEPENDENCE AND SUPERIOR PERFORMANCE ON LOCAL TASKS

As explained above, autistic LUT learning is relatively local and context-independent compared with typical INT learning. Thus, our theory agrees with the well-known Weak Central Coherence theory (Frith and Happe, 1994; Happe and Frith, 2006) and what it explains, such as autistic people's superior performance on local tasks due to reduced contextual interference. Indeed, narrow tuning of the LUT style may be a physiological cause for autism's weak central coherence. However, our theory contrasts INT learning's strong ability and LUT learning's weak ability to extract complex regularities for generalization, and as the rest of this paper shows, explains autism's poor predictive ability, surprises and over-reaction, weak adaptation and habituation, sensory overwhelm, impaired selection of one among many voices, absolute pitch, etc.; it is not immediately clear how local focus alone could account for these behaviors.

In our framework, autistic local scale results from narrow, context-independent tuning for the features under consideration, and does not always mean spatial scale. For example, for face perception, the scale is not so much determined by sizes of, or distances between, face images but by tuning "distances" in a face or facial-feature space. Thus, our theory predicts that simply reducing spatial dimensions of complex stimuli like faces will not alleviate autistic people's difficulty with such stimuli.

Although autistic people's focus on details affords them superior performances on local tasks and even helps them extract "if p, then q" type of precise, deterministic associations (Baron-Cohen, 2002), it is disadvantageous in other situations. In vision, for example, local information is often ambiguous in specifying the world, and context provides the necessary constraints in light of the statistical regularities of the world. When an experiment is designed in such a way that context interferes with a local task, autistic people show an advantage because of their weak context dependence. However, under natural conditions, context dependence is essential for statistically sound information processing and inference. Irrational fear (e.g., hearing a friend's accident produces fear of imminent danger to oneself; Robison, 2011) may also be context ignorance. Indeed, we view many of autism's such "logical" behaviors as failure to take larger, and often complex, statistical contexts into account.

IMPAIRED ATTENTIONAL SELECTION AND SWITCHING

Voluntary, top-down attentional selection of a stimulus, and switching among different stimuli, are important mechanisms that direct neural resources to the most relevant information while filtering out irrelevant one. The best known example is the cocktail party effect: typical people can engage in one conversation while ignoring the rest, and switch conversations when desired. Autistic people have difficulty with such attentional selection and switching (Courchesne et al., 1994; Grandin, 2006).

Our theory explains this impairment. A prerequisite of attentional selection and switching is to separate different sources of stimulation (e.g., voices of different people) according to their distinct statistical regularities (e.g., by maximizing independence between separated components; Bell and Sejnowski, 1995). Autistic people's LUT style cannot extract regularities, and consequently cannot separate different sources of stimulation. They have to either listen to all voices as an incomprehensible jumble or suppress them all. Their hyper-focus on a single task may help them suppress incomprehensible stimuli (Robison, 2011).

Poly-sensory cells in the brain receive inputs from different sensory modalities (Andersen et al., 1997). If high-level attention mechanisms rely on these cells but the mechanisms in autism cannot separate spiking inputs from different modalities based on their different firing statistics, then autistic people may also have difficulties with attentional selection and switching among different modalities. Moreover, sensory overload in one modality may impair attentional processing in another modality.

Another contributor to autistic people's attention deficit is their poor predictive ability (see below) because prediction is a mechanism for attention (Colby and Goldberg, 1999; Rao, 2005).

POOR PREDICTIVE ABILITY

Prediction is regularity-based generalization. For example, after observing enough moving objects, one understands momentum and can extrapolate to predict the position of a moving target in the near future. Similarly, after seeing enough human faces, one can generalize (and predict) that human faces all have similar parts in a nearly fixed spatial layout. Prediction not only affords quicker and more accurate reaction but also more efficient neural coding. For example, after learning the common face structure, the brain can store an average face and encode individual faces only as deviations from the average (Leopold et al., 2001, 2006). This is more efficient than encoding individual faces fully.

Autistic LUT style is poor at learning regularities, and therefore has poor predictive ability. In particular, narrow temporal tuning means that LUT learning operates over a short time scale, and cannot extrapolate well to predict the near future. It is better at learning correlations of events that happen at nearly the same time such as visuo-audial synchrony of clapping hands (Klin et al., 2009). Thus, our theory explains why autistic people fail to anticipate the timing of the air puff to their eyes (they blink too soon; Sears et al., 1994), and why they fail to anticipate future positions of moving objects (Sinha, 2011). A related possibility is that autistic brains may rely more on correlational Hebbian learning than predictive spike-timing-dependent plasticity (Dan and Poo, 2004) as the latter is better at learning temporal structures.

The impairment from poor prediction is not limited to the time dimension. Failure to extract regularities of faces and to use efficient predictive face coding, for example, means that autistic people have difficulty with face processing. This argument applies to any tasks and stimuli that benefit from prediction.

Kanner (1943) described two autistic children who showed signs of poor predictive ability. For one, he wrote: "The mother recalled that he was never observed to assume an anticipatory posture when she prepared to pick him up." For the other, he wrote: "The mother, in comparing her two children, recalled that while her younger [typical] child showed an active anticipatory reaction to being picked up, Richard had not shown any physiognomic or postural sign of preparedness." Social disinterest or anxiety may also contribute to these observations. However, according to our theory, autism's social difficulty itself derives from the LUT style's poor ability to learn, predict, and generalize in social situations.

Weak predictive ability might also contribute to autistic children's sometimes dangerous behaviors; perhaps they could not fully anticipate serious consequences of certain actions, such as running their heads into walls or wandering off alone. Poor prediction of bodily movements and sensory consequences of such movements might even contribute to their weak sense of body boundary and ownership (Kanner, 1943).

Another function of prediction is to fill in missing information. For example, when a saccade target disappears before the saccade, predictive reverberating activities of parietal neurons keep the target location in memory and guide the saccade to it (Gnadt and Andersen, 1988; Colby and Goldberg, 1999). Thus, autistic people's impaired predictive ability means impaired filling-in of missing information for perception and action.

HYPER- AND HYPO-SENSITIVITY, SURPRISES, ANXIETY, AND WEAK HABITUATION AND NORMALIZATION

At one moment, a touch or noise may make autistic people scream or jump; at another moment, they may not respond to calling of their names, and act as if the rest of the world did not exist (Kanner, 1943; Volkmar et al., 1986; Oneill and Jones, 1997). Such unusual reactions to sensory stimuli that are neither exceedingly strong nor weak have been referred to as hyper- and hypo-sensitivity. Hence the term "sensitivity" can be misleading because psychophysically, sensitivity refers to either the detection of weak stimuli (detection threshold) or the discrimination of very similar, but not weak, stimuli (discrimination threshold). Importantly, autistic hyper- and hypo-sensitivity, at least as originally described by (Kanner, 1943), concern neither detection of weak stimuli nor discrimination of similar stimuli; rather, they are about over and under reactions to stimuli that are typically super-detection-threshold and not subjected to fine discrimination. Therefore, although comparing detection or discrimination thresholds between autistic and typical populations is interesting in its own right (see Superior and Inferior Performance on Simple and Complex Tasks), finding differences does not fully explain hyper- or hypo-sensitivity. Indeed, standard perceptual-learning paradigms greatly reduce these thresholds in typical subjects (Gilbert, 1994; Matthews et al., 1999) without making them clinically hyper-sensitive. For example, when typical

subjects are trained to reduce their orientation discrimination thresholds drastically, they do not become over-reactive to the trained orientations.

Our explanation of autistic hyper-sensitivity relies on the simple fact that everyone can be startled by, and overreact to, *unexpected* noise, touch, etc. For typical people, unexpected stimulation is relatively infrequent and brief because their INT learning extracts regularities from sensory, motor, and social experiences, and uses these regularities to predict and anticipate roughly what will happen over various time scales. For example, they expect loud noise when seeing a hammer, but not a pillow, failing. They anticipate a hug when a friend is approaching with open arms. Even when they did not notice someone by the door, the first knock may startle them but the subsequent knocks do not because they anticipate more than one knocks. The onset of a vacuum cleaner may surprise them but they quickly predict that the noise will persist for a while and adapt to it. If it is not raining now, they do not expect rain soon. Thus, for typical people the world is reasonably predictable particularly in the near future, punctuated only by brief surprises.

In contrast, autistic LUT learning is poor at extracting regularities and thus poor at prediction and anticipation. To autistic people, then, a friendly hug may feel like a surprising squeeze, and noise from routine events may be largely unexpected and scary. Consequently, they are frequently frightened by stimuli from the world and overreact. This explains hyper-sensitivity.

This reasoning extends to most daily events. For example, typical people use prior speaking experiences to anticipate new presentations; otherwise, every presentation would be as nerve racking as the first. Thus, autistic people's weak predictive ability must make many daily events more frightening to them than to typical people. This may contribute to their high level of social anxiety or fear (Kanner, 1943; Gillott et al., 2001).

We speculate that as a defense against constant surprises from the world, as well as against overwhelming sensory stimulation and inability of attentional selection and filtering (see above), autistic people may suppress stimuli for long periods of time, possibly explaining their hypo-sensitivity.

Some anecdotes appear to support our notion that hyper-sensitivity primarily arises from poor prediction. Kanner (1943) wrote about an autistic child: "Another intrusion comes from loud noises and moving objects, which are therefore reacted to with horror… Yet it is not the noise or motion itself that is dreaded… The child himself can happily make as great a noise as any that he dreads and move objects about to his heart's desire." For another autistic child, he wrote: "He does not want me to touch him…but he will come and touch me." Our interpretation is that self-generated noise and self-initiated touch are relatively more predictable and thus less surprising. Grandin (2006) wrote that she hates hugs from people, yet, she craves touch so much that she built a machine to squeeze herself with pressure precisely controlled by herself. Again, our interpretation is that self-controlled squeeze is relatively predictable and thus no longer frightening.

Our theory further suggests that poor adaptive adjustment (normalization) of detection thresholds contributes to autistic hyper-sensitivity to constant stimuli, such as background noise in an airplane and skin pressure from clothes (Robison, 2011). Typical people adapt or habituate to such stimuli because INT learning predicts their persistence and discount them. This discounting dynamically adjusts detection thresholds to reduce or mask unimportant background stimulations. In contrast, autistic LUT style's short time scale and poor regularity learning mean weak predictive normalization of sensitivity, and consequently, weak adaptation or habituation to constant stimulations.

Our theory suggests that anxiety and fear are consequences of autistic LUT learning, rather than causes of autism. This may help resolve conflicting reports on amygdala response to faces (Baron-Cohen et al., 1999; Dalton et al., 2005). On the one hand, autistic people's poor prediction of social situations produces anxiety and fear when viewing faces, and thus increased amygdala activation. One the other hand, their poor recognition and categorization of complex stimuli such as facial emotions predicts reduced amygdala activation. The balance of these two factors may lead to either hyperactivity (Dalton et al., 2005) or hypoactivity (Baron-Cohen et al., 1999) of amygdala.

POOR PREDICTIVE CODE AND ABNORMAL BRAIN ACTIVITIES

The notion of predictive coding, which is related to redundancy reduction discussed above, posits that if a system predicts a stimulus based on statistical structure of natural world or recent experience, it only needs to encode the deviation of the actual stimuli from the prediction (Mumford, 1992; Rao and Ballard, 1999). Both low- and high-level brain areas appear to use predictive coding (Barlow and Foldiak, 1989; Leopold et al., 2006; Zhaoping, 2006). Good prediction is a form of normalization that reduces the required dynamic range of coding and consequently, the range of neuronal responses (Albrecht and Geisler, 1991; Heeger, 1992). Thus, our theory predicts that compared with typical brains, autistic brains respond more strongly to stimuli because their poor predictive ability fails to normalize neural responses. On the other hand, in perceptual filling-in phenomena, our theory predicts that cells tuned to the location of the missing inputs are more active in typical brains than in autistic brains because the former predict and fill in missing inputs whereas the latter do not.

The Imbalanced Excitation–Inhibition theory proposes that autistic brains have excess excitation, which lowers sensory detection thresholds and amplifies sensory responses (Hussman, 2001; Rubenstein and Merzenich, 2003). The theory is partly motivated by the fact that a fraction of autistic people has epilepsy and an even higher fraction has abnormal EEGs, although the causal link between epilepsy and autism is weak (Tuchman and Rapin, 2002; Levisohn, 2007). This theory and our theory differ in that we attribute abnormal brain activities to poor predictive coding of autistic LUT learning, which may lead to both increased and decreased neural activities.

PREFERENCE OF OBJECTS TO PEOPLE

Since autistic LUT style prefers precise input–output mappings, this emphasis on precision, combined with the uncertainty inflicted by their poor prediction, suggests that autistic people favor relatively predictable and precise events and tasks. Consequently, they like to play with objects, which are more predictable, instead of with people, which are less predictable particularly for autistic

people (see above). Moreover, because of the sensory overload caused by inefficient coding and poor attentional selection (see above), they must prefer objects of relatively simple and predictable configurations. On the other hand, every person, autistic or not, seeks interesting stimuli which are usually complex. Objects such as trains, bridges, and spinning wheels that autistic people are attracted to may represent interesting stimuli that their LUT-biased system can manage.

In this aspect, our theory is closely related to the Deficient Arousal Modulation theory of Dawson and Lewy (1989), who propose that autistic people have a lower tolerance to aversion induced by unpredictability and complexity. They argue that objects are predictable and simple whereas people are not, and therefore autistic people gravitate toward objects. Our theory further suggests that autistic people's aversion of unpredictability and complexity is attributable to LUT learning's insistence on precision, poor predictive ability, and inefficient coding.

SUPERIOR AND INFERIOR PERFORMANCE ON SIMPLE AND COMPLEX TASKS

Compared with controls, autistic people show superior and inferior performances on simple and complex tasks, respectively (Minshew and Goldstein, 1998; Mottron et al., 2006). For example, their discrimination of first- and second-order orientations is better and worse than controls, respectively (Bertone et al., 2005). Similarly, their processing of first-order motion may be intact, but they show deficits in processing second-order motion and biological motion (Kaiser and Shiffrar, 2009). They also have impaired recognition of facial emotions and face identity (Weeks and Hobson, 1987; Boucher and Lewis, 1992). It has been proposed that autism is a disorder in complex information processing (Minshew and Goldstein, 1998; Mottron et al., 2006), but the underlying mechanism is unclear. Our theory provides a mechanism. As we explained following Eq. 1 in the Section "Theory," narrow and broad tuning functions for autistic LUT and typical INT learning favor simple and complex tasks that require low- and high-dimensional feature spaces, respectively.

Intuitively, for a given cell, narrower tuning better specifies stimuli but covers a smaller stimulus range whereas broader tuning is less specific but covers a larger stimulus range. Equivalently, for a population of narrowly tuned cells, only a small number of them respond to a given stimulus but each cell carries a large amount of information, whereas for a population of broadly tuned cells, a large number of them respond to a given stimulus but each cell carries a small amount of information. The weighting of these factors depends on the dimensionality of the feature space in which the tuning functions are considered (Abbott and Dayan, 1999; Zhang and Sejnowski, 1999). When a d-dimensional tuning function is narrowed by a factor k in every dimension, the volume of the feature space it covers decreases by a factor k^d, and so does the number of cells contributing to a given stimulus. Consequently, for complex tasks or stimuli that require a high-dimensional feature space, the narrow tuning of autistic LUT style is disadvantageous.

Another factor is that as noted above, autistic LUT learning does not compress highly redundant input information, an inefficiency that affects complex stimuli more than simple ones. For example,

if LUT learning wastes resources to encode individual faces fully, rather than deviations from an average face, then fewer resources will be available to represent differences among faces or subtle, complex changes in a face.

Our theory predicts that autistic LUT learning prefers low-dimensional feature space because its narrow tuning functions do not benefit from high-dimensional combinatorial coding. Although formal tests are needed, an anecdote supports this prediction. Grandin (2006) wrote that she initially used size to classify cats and dogs, but after her neighbor got a small dog, she switched to using nose shape. The dimensionality of her feature space for this task appears to be one. (Her description also suggests that high-functioning autistic people can learn and generalize simple rules in a low-dimensional feature space.) Obviously, complex tasks require multiple features simultaneously (Shawe-Taylor and Cristianini, 2000). In **Figure 2**, each feature alone cannot classify the data as well as the two features combined.

Many socially relevant signals are subtle and best represented by a combination of many features. For example, a frown in a face is not a size change of a facial part but a coordinated change of many facial parts and their relative relationships. Thus, autistic people's preference of low-dimensional feature spaces must contribute to their difficulty to pick up subtle social cues.

Autistic people's superior and inferior performances on simple and complex tasks pose a challenge to many other autism theories. For example, the Imbalanced Excitation–Inhibition theory posits that excess excitation adds noise to the brain (Rubenstein and Merzenich, 2003); this predicts impaired performances for all tasks that use supra-detection-threshold stimuli (so that stochastic resonance is irrelevant (Gong et al., 2002)). On the other hand, if excess excitation boosted neuronal responses without changing tuning functions or Fano factors, then autistic people would have superior performance on all tasks because the ratio of mean response to noise SD increases with the mean response. In contrast to the notion of excess excitation, Bertone et al. (2005) argue that an excess recurrent inhibition in autism sharpens orientation tuning, which improves orientation discriminability (Regan and Beverley, 1985; Teich and Qian, 2003). However, this sharpening in one-dimensional orientation space alone cannot explain why autistic people perform worse on complex tasks.

CONCRETENESS

Various case reports suggest that autistic people think concretely instead of abstractly. Grandin (2006) wrote that whenever she hears the word "cat," she always has a vivid, detailed image of her first pet cat, and then the second pet cat, etc.; typical people are more likely to have an image of a generic or conceptual cat without much detail. Kanner (1943) wrote about an autistic child who "can set the table for a number of people if the names are given her or enumerated in any way, but she cannot set the table 'for three.'" For another autistic child, he wrote: "When asked to subtract 4 from 10, he answered 'I will draw a hexagon.'"

Our theory explains this concreteness. Recall that autistic LUT learning stores individual examples precisely without extracting underlying regularities. For example, autistic people may store the

details of specific animals they encounter together with labels "cat," "dog," etc., but they are poor at learning regularities that define cats, dogs, etc. as categories, which are regularity-based abstractions of large numbers of individual examples (**Figure 2**). Likewise, numbers are abstractions of real-world counting examples. Since the LUT style cannot learn regularities that enable abstraction, autistic people must resort to specific, concrete examples.

In contrast, typical INT learning extracts regularities from specific examples, and these regularities define abstractions such as categories and numbers. Since coding regularities/abstractions are far more efficient than coding a large number of individual examples, INT learning deemphasizes coding details of individual examples unless there is a need to do so.

Language facilitates abstract thinking and communication. Even "concrete" words such as "exhausted" are really abstractions of many related instances. Autistic people's poor ability of abstraction may contribute to their language problems and their preference of "thinking in pictures" to thinking in language (Grandin, 2006).

WEAK FACE-IDENTITY AFTEREFFECT AFTER FACE ADAPTATION
Compared with controls, autistic people show a much weaker face-identity aftereffect after face adaptation (Pellicano et al., 2007). Our theory offers a few related explanations. First, adaptation to a stimulus provides a temporal context for subsequent test stimuli, and in this sense, aftereffects are contextual effects. The LUT style's narrow temporal tuning limits temporal context and weakens aftereffects. Second, for controls the largest aftereffects usually occur for stimuli somewhat different from the adaptor. For example, orientation tilt aftereffect is largest for test orientations about 15° away from the adapting orientation (Mitchell and Muir, 1976), which can be explained by typical orientation tuning widths (Teich and Qian, 2003).When the tuning is narrower, however, the largest aftereffects will be not only closer to the adapting stimulus, but also smaller in magnitude because narrow tuning limits peak shifts of neuronal population responses, which are responsible for aftereffects. Therefore, narrow face tuning of autistic LUT learning may contribute to the small face-identity aftereffect.

Third, as mentioned before, face processing requires a high-dimensional feature space. Consequently, normal face-identity aftereffect may result from adaptation in multiple dimensions of feature space. Since narrow tuning of autistic LUT learning favors a low-dimensional feature space (see Theory), fewer dimensions contribute to autistic face-identity adaptation and hence a smaller aftereffect.

It has been shown that there are both local and holistic components in face representation (Xu et al., 2008). The former concerns individual face parts whereas the latter involves non-linear combinations of the parts. The finding of autistic people's impaired holistic face processing (Joseph and Tanaka, 2003) suggests that they may indeed use a very low-dimensional feature space that focuses on one or two face parts. This predicts that adaptation to relevant face parts should account for most of their face adaptation aftereffect.

Finally, prediction is a mechanism for adaptation (Grzywacz and de Juan, 2003). Thus, autistic people's poor predictive ability implies reduced adaptation and habituation (see above).

ABSOLUTE PITCH AND ABSOLUTE VS. RELATIVE JUDGMENTS
We are much better at making relative judgments (e.g., point A is closer than point B) than absolute judgments (e.g., point A is 5.12 m away). The reason may be that relative relationships better reflect useful regularities of the world. For example, the absolute distance of an object varies greatly with the observer's locomotion whereas the relative depth orders between parts of an object are more invariant and define the object shape. Similarly, absolute pitch, pace, and loudness of a person's speech does not carry as much information as modulations relative to the means (prosody), which convey emotion, emphasis, sarcasm, etc. Thus, typical brains must devote more resources to encode relative than absolute quantities. It is also more efficient to encode small, relative values than to encode large, absolute values (analogous to the earlier discussion of predictive face coding).

Autistic LUT style is poor at learning regularities, and thus may not emphasize relative over absolute coding as much as typical INT style. Consequently, autistic people must have a stronger tendency to code absolute quantities, which explains their better absolute pitch ability (Heaton et al., 1998). Also note that pitch is a simple feature and narrow tuning of autistic LUT style may have an advantage in coding it precisely (see above). As for relative judgments, autistic people should do better than controls for simple stimuli by comparing precise absolute estimations, but worse for complex stimuli which overload their inefficient system and are not well represented by their low-dimensional feature spaces.

In the limit of very narrow tuning, cells act as labeled lines for the narrow range of stimulation they respond to. Therefore, activation of a specific set of cells directly reports the absolute value of stimulation. Broad tuning functions can also make absolute estimations. However, with broad tuning, a stimulus activates a response distribution in many cells and additional steps (such as Bayesian inference or maximum likelihood estimation) are needed to compute absolute values. Thus, absolute judgments may be easier with narrow than with broad tuning.

EXPERIMENTAL TESTS OF OUR THEORY
The theory's most direct prediction is that autistic and typical brains are biased toward LUT and INT styles of learning, respectively. This can be tested by training subjects on random (but fixed) association tasks and tasks with hidden, underlying rules. We predict that compared with age- and IQ-matched controls, autistic people do better on the former tasks but worse on the latter tasks particularly when the rules are complex or noisy. Importantly, this prediction does not depend on whether or not the tasks have social relevance. In fact, it is best to use non-social tasks (e.g., learning visual categorization of shapes) to avoid potential confounds from autistic and typical subjects' different developmental and intervention histories.

Another test concerns coding efficiency. After learning complex hidden rules for a task, controls can apply the rules rather than memorize individual examples. Autistic people, however, cannot learn complex rules well and try to store specific examples. We thus predict that increasing the number of examples will not affect controls (after they have learned the rules) but will increasingly burden autistic subjects and slow their learning.

Our theory hypothesizes that autistic LUT learning has a stronger emphasis on storing training data precisely than does typical INT learning. This can be tested in the random association tasks above by asking subjects to reproduce the training examples, and examine whether autistic subjects are more precise than controls.

Our theory predicts that typical INT and autistic LUT learning lead to strong and weak adaptation aftereffects, respectively. Pellicano et al.'s (2007) face experiment already supports this prediction but the result should hold for other stimuli instead of face specific. A related prediction is that autistic people show weak adaptive normalization of detection thresholds as a function of background stimulation because of their impaired predictive coding.

SPECTRUM NATURE OF AUTISM
Our theory is consistent with the spectrum nature of autism because the degree of LUT learning may vary among different individuals and among different modalities/systems/tasks for the same individual. We mentioned that different systems of the brain, autistic or not, may have different biases toward INT and LUT learning depending on what kind of tasks the system does. This natural variation of learning styles among different systems suggests that autism may affect different systems to varying degrees. Although we often contrast typical vs. autistic populations for ease of description, our theory suggests a continuum. In particular, high-functioning autistic people must be able to extract certain regularities from experience by using INT learning to some extent [e.g., Robison (2011) can attend to sound of a particular instrument in a concert]. Typical people may also have a few autistic traits without meeting diagnostic criteria and some of those traits (e.g., strong focus, attention to detail) are advantageous.

DISCUSSION
SUMMARY OF OUR THEORY AND EXPLANATIONS
Unlike previous efforts that focus on small subsets of autistic behaviors, we propose a theory that appears to account for a wide range of autistic behaviors. We hypothesize that autistic brains are biased toward LUT learning, which aims to store training examples precisely without extracting their underlying statistical structure or regularities, whereas typical brains are biased toward INT learning, which does not insist on representing training examples precisely but focuses instead on discovering their underlying regularities for generalization. These learning styles may be implemented by relatively narrow and broad tuning functions and strong and weak emphasis on eliminating small errors during learning, respectively. The narrow and broad tuning functions also imply that LUT and INT learning favor low- and high-dimensional feature spaces, respectively.

The LUT style is good for learning relationships that are local, precise, rigid, and do not contain inherent regularities for generalization, such as name–number association in a phonebook. However, it is poor at learning relationships that are context dependent, noisy, flexible, and do contain regularities for generalization. Since most relationships in social interaction, language/communication, and sensorimotor processing are of the latter type, our theory explains autism's broad range of behaviors. In particular, autistic people's preference of non-social to social stimuli (Osterling and Dawson, 1994; Klin et al., 2009) reflects what they are able to learn, rather

than a lack of social interest, *per se*. Because of their poor regularity learning and generalization, but good memorization of specific examples, autistic people rely on recalling the memory entry that best matches the current situation, resulting in rigid social behavior and formulaic language (Kanner, 1943). Similarly, without learning regularities that define categories and concepts abstractly, autistic people rely on specific examples and appear concrete (Kanner, 1943; Grandin, 2006). Reduced generalization also explains their performances, including some superior ones, in word-list (Beversdorf et al., 2000), perceptual (Plaisted et al., 1998), and motor learning tasks (Haswell et al., 2009).

Context dependence in sensorimotor processing, language, and social interaction arises from statistical regularities in these domains. Thus, autistic people's poor regularity learning impairs their context dependence. Their narrow tuning and consequent low-dimensional feature space further limit their ability to cover large scales or use extra dimensions to represent context. Context independence leads to stereotyped social behavior, literal interpretation of language, and poor sensory processing and motor control. It also leads to superior performance of local tasks because of reduced contextual interference (Frith and Happe, 1994; Happe and Frith, 2006). Autistic LUT learning's low-dimensional feature spaces also explain their poor performances on complex stimuli and tasks that require more feature dimensions (Minshew and Goldstein, 1998; Mottron et al., 2006).

By ignoring regularities in training examples, LUT learning stores information in relatively raw forms without adequate compression to remove redundancy, and is thus inefficient, resource intensive, and easily overwhelmed by information overload (Kanner, 1943; Grandin, 2006; Robison, 2011). Consequently, given fixed neuronal resources, the algorithm can only learn a restricted set of behavioral repertoires. Moreover, the algorithm insists on learning noisy mappings precisely, and this over training on a limited set of behaviors makes the system harder to break acquired habits to learn new ones. Additionally, a system with LUT learning cannot easily select one among many sources of stimulation (Courchesne et al., 1994; Grandin, 2006) because a prerequisite of such attentional selection is to separate different sources according to their distinct statistical regularities. Furthermore, since prediction is regularity-based generalization, LUT learning implies poor predictive ability, resulting in surprises, over-reaction (hyper-sensitivity), anxiety, and fear (Kanner, 1943; Volkmar et al., 1986; Oneill and Jones, 1997; Gillott et al., 2001). As a defense against surprises, as well as against sensory overwhelm and impaired attentional selection, autistic people may suppress stimulation (hypo-sensitivity) and prefer relatively predictable situations (Dawson and Lewy, 1989).

The spectrum nature of autism can be explained by different degrees of LUT learning among different individuals and in different dimensions of the same individual. High-functioning autistic people must be able to learn and generalize relatively precise, low-dimensional regularities using limited INT learning.

RELATIONSHIPS BETWEEN OUR THEORY AND PREVIOUS THEORIES
Our theory is related to, but distinct from, previous theories. For example, LUT learning's local focus is related to the Weak Central Coherence theory (Frith and Happe, 1994; Happe and

Frith, 2006). LUT learning's impaired predictive ability is related to the Deficient Arousal Modulation theory (Dawson and Lewy, 1989). LUT learning's poor attentional selection is related to the Attention Deficit theory (Courchesne et al., 1994). LUT learning's superior and inferior performance on simple and complex tasks is related to the Complex Information Processing Disorder theory (Minshew and Goldstein, 1998; Mottron et al., 2006). And LUT learning's insistence on precision is related to the Over-Fitting theory (Cohen, 1994). Importantly, our theory suggests an underlying framework that unifies and encompasses many existing theories (including, but not limited to, the ones mentioned above) by combining the strengths of those theories without concatenating their different assumptions. Rather, those theories, and other implications, are logical consequences of our root assumption of different learning styles for autistic and typical populations. Therefore, our theory explains more autistic behaviors without making more assumptions. Although each of the phenomena listed in the Introduction has probably been explained before, to our knowledge, they have not been explained together coherently by a single theory. Additionally, our explanations of some autistic behaviors (e.g., sensory overwhelm, hyper-sensitivity, restricted interests, atypical learning, impaired attentional selection, concreteness, weak adaptation, absolute pitch, and inferior performance on complex tasks) differ from previous theories. Finally, our testable predictions on learning style and efficiency of autistic and typical subjects (see above) are not made by previous theories.

One may argue that some autistic behaviors we discussed can be alternatively explained by autism's social disinterest and anxiety. For example, high-functional autistic people can use size to (inadequately) classify cats and dogs (Grandin, 2006) and use numbers 1, 2, and 3 to name three sisters (Robison, 2011) but fail to pick up social cues. According to our theory, this is because their limited INT learning allows them to extract simple, precise rules for generalization (in low-dimensional feature spaces) but has trouble with complex, context-dependent regularities (in high-dimensional feature spaces) that are typically found in social situations. The alternative argument is that autism only has a generalization problem in social settings but generalizes fine otherwise. However, this alternative does not address the issue of why autism has social difficulty in the first place. Our theory avoids this chicken-and-egg problem as we explain all autistic behaviors, including social difficulty, with the assumption that autistic brains' LUT bias reduces their ability to extract complex regularities for generalization and prediction, and enhances their ability to memorize examples. Possible physiological and anatomical basis for the LUT style is discussed below. Also note that our and the alternative explanations can be distinguished by using non-social learning tasks with complex, noisy rules (see Experimental Tests of Our Theory).

We compare our theory with three additional theories. The first is the Inadequate Cortical Feature Map theory (Gustafsson, 1997) which assumes that autism involves excessive lateral inhibition, leading to smaller cortical columns and defective feature maps. Our theory similarly predicts that autistic people have difficulty extracting complex features. However, our root assumption of LUT vs. INT learning is not part of the Inadequate Cortical

Feature Map theory (or other theories). Consequently, our theory explains many autistic behaviors differently and accounts for more behaviors. For example, our theory implies that typical people can learn complex regularities in high-dimensional feature spaces whereas autistic people prefer simple, precise rules in low-dimensional feature spaces. This explains, among other things, autistic people's superior and inferior discriminability for simple and complex stimuli, respectively, compared with typical people. The Inadequate Cortical Feature Map theory does not make this argument but appears to suggest that smaller cortical columns lead to both generally superior feature discriminability and impaired feature extraction. Our theory explains autism's strong rote memorization as reduced interference between new and old memories due to LUT learning's narrow tuning functions. In contrast, The Inadequate Cortical Feature Map theory argues that impaired feature extraction forces rote memory. We attribute autism's sensory overwhelm and restricted interests to LUT learning's weak information compression. Instead, the Inadequate Cortical Feature Map theory suggests that an "autistic individual might choose narrow interests simply because he would not have the capacity to deal with several interests in the same consuming manner." We explain autistic people's overreaction and reduced adaptation to sensory stimuli as a consequence of LUT learning's poor predictive ability, and their impaired attentional selection and switching as resulting from LUT learning's inability to separate multiple sources of stimulation. It is not clear how the Inadequate Cortical Feature Map theory explains these behaviors.

The second theory is the Intense World theory (Markram and Markram, 2010) which, like our theory, aims at a unified account of autism. Based on an animal model, the Intense World theory posits that autistic "neuropathology is hyper-functioning of local neural microcircuits, best characterized by hyper-reactivity and hyper-plasticity." It is then believed that this leads to "dominance of the earliest features" and "avoidance of processing of other features." However, the theory further argues that "the lack of social interaction in autism may therefore not be because of deficits in the ability to process social and emotional cues, but because a subset of cues are overly intense[…]". In contrast, our theory suggests that autistic LUT learning is poor at extracting subtle, complex social cues. The Intense World theory also argues that hyper-plastic systems could "become autonomous and memory trapped." However, computationally, hyper-plasticity implies generally fast learning, fast forgetting, and weak convergence, and thus highly fluid, instead of trapped, systems. Trapping could occur under other conditions, e.g., when impaired synaptic normalization or homeostasis fails to rescue saturated synapses, and hyper-plasticity may accelerate the process to saturation. However, once synapses are stuck in saturation, they are no longer plastic, let alone hyper-plastic, and if synapses are hyper-plastic, they cannot be stuck. Therefore, it is not clear to us how the Intense World theory explains many of the autistic behaviors that our theory accounts for coherently.

The third theory is the well-known Extreme Male Brain theory (Baron-Cohen, 2002) which posits that female and male brains prefer empathizing and systemizing, respectively, and that autism is an extreme form of the male brain (hyper-systemizing). Here, "system-

izing" means extracting "if p, then q" type of rules from "systems" defined by such rules. The social world is not a "system" according to this theory and is deemed understandable only through empathizing. This dichotomy between systems and non-systems may roughly correspond to our contrast between rigid, precise relationships and noisy, flexible relationships which are better learned by the LUT and INT styles, respectively. However, we do not directly use the concept of empathizing; instead, we believe that the social world also has underlying (albeit fuzzy and context dependent) rules that can be extracted from experiences via typical INT learning and that empathy may be viewed as a mental switch of context. Autism's apparent lack of empathy (Baron-Cohen et al., 1985; Baron-Cohen, 2002) may thus be attributable to LUT learning's context deficit. Moreover, we posit that autistic LUT bias is poor at extracting and generalizing complex rules in high-dimensional feature spaces but can readily store "if p, then q" type of simple, precise rules as an association. Hyper-systemizing may then correspond to autistic people's application of simple, precise rules that they manage to learn [e.g., Robison's (2011) naming of his wife and her sisters as units 1, 2, and 3]. Finally, we view autism's repetitive behaviors as a consequence of over training on limited behavioral repertoires constrained by the inefficient LUT style whereas the Extreme Male Brain theory views repetitions as systemizing efforts.

Many brain areas (amygdala, hippocampus/limbic system, frontal/prefrontal cortex, parietal cortex, cerebellum, basal ganglion, fusiform face area, superior temporal sulcus, mirror neuron system, locus coeruleus, etc.) have been implicated in autism theories and experiments, suggesting that no single area dictates the disorder. We speculate that abnormalities found in an area reflect mutual interactions between abnormal learning/development and abnormal structure/physiology in that area.

ASSUMPTIONS AND POSSIBLE NEURAL MECHANISMS

Our key assumption is the LUT and INT learning styles for autistic and typical brains, respectively. We argued that different tuning widths help realize these different learning styles (**Figure 1**) and also lead to different dimensionalities of the feature space for learning (Eq. 1). Thus, the tuning-width assumption parsimoniously combines learning style and feature-space dimensionality, which are essential for the explanatory power of our theory. The tuning-width assumption also provides a possible converging point for the actions of diverse autism genes and perhaps multiple anatomical/physiological substrates.

Our framework is at the functional or computational level (Marr, 1982). Physiologically, we interpret tuning generally to include contributions from both classical receptive field (CRF) and extra-classical receptive field (ECRF or surround) which modulates CRF responses (Allman et al., 1985). The CRF arises from feedforward connections from lower areas and short-range interactions among neighboring cells, and encodes local stimulus features (Hubel and Wiesel, 1962). The ECRF arise from both long-range horizontal connections within an area (Gilbert and Wiesel, 1990) and feedback connections from higher areas (Angelucci and Bressloff, 2006), and is important in interpreting local features in each CRF according to the context provided by both the bottom-up stimulations over a large area (Gilbert and Wiesel, 1990; Zhaoping, 2006) and top-down predictions based

on experienced statistical regularities (Mumford, 1992; Rao and Ballard, 1999). Since functionally, autistic LUT learning shows a strong local focus but poor context dependence and prediction, a possible physiological mechanism for LUT learning is diminished surround, ECRF influences on strong, local CRF responses (Ken Miller, personal communication). Anatomically, this implies reduced long-range connections within an area, or reduced feedback connections from higher areas, or both, to weaken ECRF, and enhanced short-range connections among neighboring cells, or enhanced feedforward connections from lower areas, or both, to strengthen CRF. Thus, this possibility might relate our theory to the Under Connectivity theory (Casanova et al., 2002; Just et al., 2004; Courchesne and Pierce, 2005; Tommerdahl et al., 2008) although that theory does not distinguish feedback and feedforward connections between areas. The proposed anatomical changes also suggest enhanced stimulus selectivity within CRF and reduced stimulus range in ECRF that modulates CRF responses, consistent with the assumed narrow tuning of the LUT style (**Figure 1**).

The above scenario of relatively isolated local cell groups raises a closely related possibility for implementing the LUT style. If an autistic brain recruits a small group of cells to store each training example and if there is little overlap or connectivity among different cell groups, then there will be little interference among stored examples and little interpolation across examples for regularity extraction, as required by LUT learning. Moreover, the number of cells in each group has to be small and the number of groups has to be large in order for the system to store a large number of examples. This then implies a low-dimensional feature space (again as required by LUT learning) since a joint, combinatorial representation of many features would require a large number of cells in each group. This possibility appears to be broadly consistent with the Inadequate Cortical Feature Map theory (Gustafsson, 1997) and anatomical and physiological evidence for smaller and more numerous mini-columns in autism (Casanova et al., 2002; Tommerdahl et al., 2008). Both proposed implementations above assume that the LUT vs. INT learning styles are realized at the level of local cell assemblies without excluding cellular-level mechanisms (e.g., different synaptic plasticity rules).

Although the above discussion suggests potential links between our theory and the Under Connectivity theory, it does not make these theories identical. First, our core assumption of LUT vs. INT learning, which is responsible for the explanatory power of our theory, is not included in any previous theories. Second, the Under Connectivity theory is consistent with, but does not logically imply, narrow tuning or LUT learning. In fact, reducing or increasing any connections could either broaden or sharpen tuning, or make little difference, depending on the nature of the connections. For example, reducing well-aligned connections from LGN to V1 can broaden orientation tuning whereas reducing mis-aligned connections can sharpen the tuning; either effect becomes negligible if V1 cells form a recurrent attractor network (Teich and Qian, 2006). Even when previous studies mention narrow tuning, the discussions were never about LUT learning. Third, whether the circuits outlined above could really implement LUT learning or whether there are other implementations are open questions that require further

investigation. In this aspect, our functional-level theory could have the advantage of unifying multiple implementations. Fourth, we suggested above a distinction between feedforward connections from low- to high-level areas and feedback connections from high- to low-level areas whereas the Under Connectivity theory does not. Finally, to our knowledge, the Under Connectivity theory has not been applied to explain all of the autistic behaviors that we have considered in this paper.

THERAPEUTIC IMPLICATIONS

Our theory suggests that the difficulty of training autistic people to learn typical behaviors is the same as the difficulty of training typical people to memorize random factual details such as phone numbers; both arise from a mismatch between learning style and task. Since most relationships in social interaction, language, and sensorimotor processing contain fuzzy, flexible, and complex regularities, therapies should focus on how to train autistic LUT style to learn these regularities. We suggest a few possibilities.

First, there are mnemonic tricks that help typical people to remember random facts such as phone numbers or digits in π. The idea is to associate random facts with a coherent theme or story that is easy for typical INT style to learn and remember. Perhaps, those tricks could be reversed to help autistic people. The key empirical question is whether we can train autistic people to use lists of memorized random facts to code regularities in social interaction, language, and even sensorimotor processing.

Second, high-functioning autistic people clearly have some ability to learn certain regularities. It might be possible to extend their ability through systematic training. Since they have difficulty discovering regularities, particularly complex ones, regularities should be explained to them explicitly. It might also help to start training them on simple regularities and gradually move on to more complex ones. For example, if they have trouble separating different sound sources, then training should start with a mixture of two very different sounds that can be distinguished in one dimension, say, loudness. Then they can be trained on sounds that require a two-dimensional feature space, say, loudness and pitch, to separate, and so on.

Third, autistic people could be explicitly instructed on why learning a general regularity is more useful than storing specific examples precisely. They could be explicitly taught how to apply regularity to perform a task, such as classifying cats and dogs, without memorizing individual examples.

Finally, if future experiments could confirm that narrow tuning is indeed a property of autistic brains, then any drug or manipulation that could broaden tuning functions would help.

LIMITATIONS OF OUR THEORY

First, the theory so far is qualitative. Quantitative simulations will pose a major challenge, particularly for complex tasks such as language and social interaction, due to the lack of mechanistic neural models. Second, our theory is largely at the computational level (Marr, 1982). We discussed a possible neural mechanism, but much work is needed to specify the detailed micro-circuitry and synaptic plasticity rules that implement LUT and INT learning. Third, our theory cannot explain why the male-to-female ratio in autistic population is about 4–1, unless there were links of gender to learning style. The Extreme Male theory (Baron-Cohen, 2002) explains this result but how it may account for many of the autistic behaviors listed in Introduction is unclear. Finally, our theory cannot explain why fever relieves autistic symptoms unless fever could change learning style. The Locus Coeruleus theory (Mehler and Purpura, 2009) raises an intriguing connection between fever and autism but is not specific about autistic behaviors.

In summary, we have proposed a learning-style theory that accounts for a wide range of autistic behaviors. The theory needs further quantification and grounding on physiology and anatomy. Most importantly, its key assumptions and implications need to be tested experimentally.

ACKNOWLEDGMENT

REFERENCES

Abbott, L. F., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11, 91–101.

Albrecht, D. G., and Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis. Neurosci.* 7, 531–546.

Allman, J., Miezin, F., and McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu. Rev. Neurosci.* 8, 407–430.

Andersen, R. A., Snyder, L. H., Bradley, D. C., and Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* 20, 303–330.

Angelucci, A., and Bressloff, P. C. (2006). "Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons," in *Progress in Brain Research*, eds S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, and P. U. Tse (Amsterdam: Elsevier), 93–120.

Barlow, H., and Foldiak, P. (1989). "Adaptation and decorrelation in the cortex," in *The Computing Neuron*, Chap. 4, eds R. Durbin, C. Miall, and G. Mitchinson (New York: Addison-Wesley), 54–72.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cogn. Sci. (Regul. Ed.)* 6, 248–254.

Baron-Cohen, S., Ashwin, E., Ashwin, C., Tavassoli, T., and Chakrabarti, B. (2009). Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1377–1383.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., and Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *Eur. J. Neurosci.* 11, 1891–1898.

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159.

Bertone, A., Mottron, L., Jelenic, P., and Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain* 128, 2430–2441.

Beversdorf, D. Q., Smith, B. W., Crucian, G. P., Anderson, J. M., Keillor, J. M., Barrett, A. M., Hughes, J. D., Felopulos, G. J., Bauman, M. L., Nadeau, S. E., and Heilman, K. M. (2000). Increased discrimination of "false memories" in autism spectrum disorder. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8734–8737.

Boucher, J., and Lewis, V. (1992). Unfamiliar face recognition in relatively able autistic children. *J. Child Psychol. Psychiatry* 33, 843–859.

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

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

Colby, C. L., and Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annu. Rev. Neurosci.* 22, 319–349.

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

Courchesne, E., Townsend, J., Akshoomoff, N. A., Saitoh, O., Yeungcourchesne, R., Lincoln, A. J., James, H. E., Haas, R. H., Schreibman, L., and Lau, L. (1994). Impairment in shifting attention in autistic and cerebellar patients. *Behav. Neurosci.* 108, 848–865.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., and Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* 8, 519–526.

Dan, Y., and Poo, M.-M. (2004). Spike timing-dependent plasticity of neural circuits. *Neuron* 44, 23–30.

Darian, S. (1979). The role of redundancy in language and language teaching. *System* 7, 47–59.

Dawson, G., and Lewy, A. (1989). "Arousal, attention, and the socioemotional impairments of individuals with autism," in *Autism: Nature, Diagnosis, and Treatment*, ed. G. Dawson (New York: Guilford Press), 49–74.

Dawson, M., Mottron, L., and Gernsbacher, M. A. (2008). "Learning in autism," in *Learning and Memory: A Comprehensive Reference III Cognitive Psychology*, eds J. Byrne and H. L. Roediger (Oxford: Elsevier), 759–772.

Diehl, J., Bennetto, L., and Young, E. (2006). Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *J. Abnorm. Child Psychol.* 34, 83–98.

Frith, U., and Happe, F. (1994). Autism – beyond theory of mind. *Cognition* 50, 115–132.

Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Netw.* 2, 183–192.

Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* 41, 711–724.

Gilbert, C. D. (1994). Early perceptual learning. *Proc. Natl. Acad. Sci. U.S.A.* 91, 1195–1197. [comment].

Gilbert, C. D., and Wiesel, T. N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Res.* 30, 1689–1701.

Gillott, A., Furniss, F., and Walter, A. (2001). Anxiety in high-functioning children with autism. *Autism* 5, 277–286.

Girosi, F., and Poggio, T. (1990). Networks and the best approximation property. *Biol. Cybern.* 63, 169–176.

Gnadt, J. W., and Andersen, R. A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* 70, 216–220.

Gong, Y., Matthews, N., and Qian, N. (2002). Model for stochastic-resonance-type behavior in sensory perception. *Phys. Rev. E* 65, 031904.

Grandin, T. (2006). *Thinking in Pictures: And Other Reports from My Life with Autism*. New York: Vintage Books.

Grzywacz, N. M., and de Juan, J. (2003). Sensory adaptation as Kalman filtering: theory and illustration with contrast adaptation. *Network* 14, 465–482.

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

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

Haswell, C. C., Izawa, J., Dowell, L. R., Mostofsky, S. H., and Shadmehr, R. (2009). Representation of internal models of action in the autistic brain. *Nat. Neurosci.* 12, 970–972.

Heaton, P., Hermelin, B., and Pring, L. (1998). Autism and pitch processing: a precursor for savant musical ability? *Music Percept.* 15, 291–305.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9, 181–197.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* 160, 106–154.

Hussman, J. (2001). Letters to the editor: suppressed GABAergic inhibition as a common factor in suspected etiologies of autism. *J. Autism Dev. Disord.* 31, 247–248.

Joseph, R. M., and Tanaka, J. (2003). Holistic and part-based face recognition in children with autism. *J. Child Psychol. Psychiatry* 44, 529–542.

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

Kaiser, M. D., and Shiffrar, M. (2009). The visual perception of motion by observers with autism spectrum disorders: a review and synthesis. *Psychon. Bull. Rev.* 16, 761–777.

Kanner, L. (1943). Autistic disturbances of affctive contact. *Nerv. Child* 2, 217–250.

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., and Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459, 257–261.

Leopold, D. A., Bondar, I. V., and Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature* 442, 572–575.

Leopold, D. A., O'Toole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nat. Neurosci.* 4, 89–94.

Levisohn, P. M. (2007). The autism-epilepsy connection. *Epilepsia* 48, 33–35.

Li, Z. (2000). Pre-attentive segmentation in the primary visual cortex. *Spat. Vis.* 13, 25–50.

Losh, M., and Capps, L. (2003). Narrative ability in high-functioning children with autism or Asperger's syndrome. *J. Autism Dev. Disord.* 33, 239–251.

Markram, K., and Markram, H. (2010). The intense world theory – a unifying theory of the neurobiology of autism. *Front. Hum. Neurosci.* 4:224. doi: 10.3389/fnhum.2010.00224

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.

Matthews, N., Liu, Z., Geesaman, B. J., and Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Res.* 39, 3692–3701.

Mehler, M. F., and Purpura, D. P. (2009). Autism, fever, epigenetics and the locus coeruleus. *Brain Res. Rev.* 59, 388–392.

Minshew, N. J., and Goldstein, G. (1998). Autism as a disorder of complex information processing. *Ment. Retard Dev. Disabil. Res. Rev.* 4, 129–136.

Minsky, M. (1985). *The Society of Mind*. New York: Simon & Schuster, Inc.

Mitchell, D. E., and Muir, D. W. (1976). Does the tilt after-effect occur in the oblique meridian? *Vision Res.* 16, 609–613.

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

Mumford, D. (1992). On the computational architecture of the neocortex. *Biol. Cybern.* 66, 241–251.

Oneill, M., and Jones, R. S. P. (1997). Sensory-perceptual abnormalities in autism: a case for more research? *J. Autism Dev. Disord.* 27, 283–293.

Osterling, J., and Dawson, G. (1994). Early recognition of children with autism: a study of first birthday home videotapes. *J. Autism Dev. Disord.* 24, 247–257.

Pellicano, E., Jeffery, L., Burr, D., and Rhodes, G. (2007). Abnormal adaptive face-coding mechanisms in children with autism spectrum disorder. *Curr. Biol.* 17, 1508–1512.

Pelphrey, K. A., Morris, J. P., and McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain* 128, 1038–1048.

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

Poggio, T., and Girosi, F. (1990). Networks for approximation and learning. *Proc. IEEE* 78, 1481–1497.

Rao, R. P. (2005). Bayesian inference and attentional modulation in the visual cortex. *Neuroreport* 16, 1843–1848.

Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87.

Regan, D., and Beverley, K. I. (1985). Postadaptation orientation discrimination. *J. Opt. Soc. Am. A* 2, 147–155.

Robison, J. E. (2011). *Be Different*. New York: Crown Archetype.

Roweis, S. T., and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 2323–2326.

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

Sears, L. L., Finn, P. R., and Steinmetz, J. E. (1994). Abnormal classical eye-blink conditioning in autism. *J. Autism Dev. Disord.* 24, 737–751.

Shawe-Taylor, J., and Cristianini, N. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Methods*. Cambridge: Cambridge University Press.

Sigman, M., Cecchi, G. A., Gilbert, C. D., and Magnasco, M. O. (2001). On a common circle: natural scenes and Gestalt rules. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1935–1940.

Sinha, P. (2011). Learning to see later in life. *Featured Lecture at 2011 Society for Neuroscience Annual Meeting*, San Diego.

Teich, A. F., and Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. *J. Neurophysiol.* 89, 2086–2100.

Teich, A. F., and Qian, N. (2006). Comparison among some models of orientation selectivity. *J. Neurophysiol.* 96, 404–419.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323.

Tommerdahl, M., Tannan, V., Holden, J., and Baranek, G. (2008). Absence of stimulus-driven synchronization effects on sensory perception in autism: evidence for local underconnectivity? *Behav. Brain Funct.* 4, 19.

Tuchman, R., and Rapin, I. (2002). Epilepsy in autism. *Lancet Neurol.* 1, 352–358.

Volkmar, F. R., Cohen, D., and Paul, R. (1986). An evaluation of DSM-III criteria for infantile autism. *J. Am. Acad. Child Adolesc. Psychiatry* 25, 190–197.

Walenski, M., Mostofsky, S. H., Gidley-Larson, J. C., and Ullman, M. T. (2008). Brief report: enhanced picture naming in autism. *J. Autism Dev. Disord.* 38, 1395–1399.

Walesski, M., Tager-Flusberg, H., and Ullman, M. T. (2006). "Language in autism," in *Understanding Autism: From Basic Neuroscience to Treatment,* eds S. O. Moldin and J. L. R. Rubenstein (Boca Raton, FL: Taylor and Francis Books), 175–203.

Weeks, S. J., and Hobson, R. P. (1987). The salience of facial expression for autistic children. *J. Child Psychol. Psychiatry* 28, 137–152.

Werthermer, M. (1938). *Laws of Organization in Perceptual Forms.* London: Harcourt, Brace, and Jovanovitch.

Xu, H., Dayan, P., Lipkin, R. M., and Qian, N. (2008). Adaptation across the cortical hierarchy: low-level curve adaptation affects high-level facial-expression judgments. *J. Neurosci.* 28, 3374–3383.

Zhang, K. C., and Sejnowski, T. J. (1999). Neuronal tuning: to sharpen or broaden? *Neural Comput.* 11, 75–84.

Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Network* 17, 301–334.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.