NEURAL DYNAMICS OF AUTISTIC BEHAVIORS:
Cognitive, Emotional, and Timing Substrates

Stephen Grossberg* and Don Seidman**
*Center for Adaptive Systems
and
Department of Cognitive and Neural Systems
677 Beacon Street
Boston University
Boston, MA 02215
steve@bu.edu
617-353-7858 (Phone)
617-353-7755 (Fax)
http://www.cns.bu.edu/Profiles/Grossberg

**DuPage Medical Group
103 Haven Road
Elmhurst, IL 60126

Corresponding Author: Stephen Grossberg

Submitted: August 2004

Running Title: Neural Dynamics of Autistic Behaviors

Key words: autism, learning, categorization, depression, hypervigilance, adaptive resonance theory, adaptive timing, amygdala, frontal cortex, hippocampus, cerebellum

ABSTRACT

What brain mechanisms underlie autism and how do they give rise to autistic behavioral symptoms? This article describes a neural model, called the iSTART model, which proposes how cognitive, emotional, timing, and motor processes that involve brain regions like prefrontal and temporal cortex, amygdala, hippocampus, and cerebellum may interact together to create and perpetuate autistic symptoms. These model processes were originally developed to explain data concerning how the brain controls normal behaviors. The iSTART model shows how autistic behavioral symptoms may arise from prescribed breakdowns in these brain processes, notably a combination of underaroused emotional depression in the amygdala and related affective brain regions, learning of hyperspecific recognition categories in temporal and prefrontal cortices, and breakdowns of adaptively timed attentional and motor circuits in the hippocampal system and cerebellum. The model clarifies how malfunctions in a subset of these mechanisms can, though a system-wide vicious circle of environmentally mediated feedback, cause and maintain problems with them all.

INTRODUCTION

1. Overview

Autism is a complex developmental disorder of pervasively distorted development. Social and communication abilities are especially affected. Kanner (1943) suggested some inborn defect in his initial report. Since that time, molecular genetics, neurochemistry, neuropathology, embryology, neurophysiology, and various different schools of psychological analysis have all contributed significantly to understanding autism. A neurally-based theoretical framework that integrates these various lines of research is, however, still lacking, despite the existence of several suggestive cognitive theories that each help to rationalize a subset of autistic behaviors (Frith and Hill, 2003).

This article describes a neural network model, called the Imbalanced START (iSTART) model, whose properties clarify possible brain mechanisms of autism and how they give rise to autistic behavioral symptoms. The model includes interactions between cognitive, emotional, timing, and motor mechanisms, and is consistent with data from a variety of disciplines that implicate early onset dysfunction of the limbic and cerebellar systems in autism. The START (Spectrally Timed Adaptive Resonance Theory) model of normal cognitive and cognitive-emotional behavior was derived over a period of years to explain many data about the brain mechanisms that control normal cognitive, emotional, timing, and motor behaviors. The iSTART model clarifies how autism may arise from prescribed breakdowns in these mechanisms. The model hereby provides a unifying interdisciplinary perspective that links normal to autistic behaviors, and embodies a number of predictions about mechanisms underlying autism which may help to integrate research from diverse fields.

This article illustrates how a complex mental disorder like autism can provide a unifying perspective for understanding many brain mechanisms, how they work together, and how specific breakdowns in these mechanisms can lead to clinical symptoms. The article can thus be read both as a review of several models that have successfully explained different types of data about normal brain and behavior, as well as a proposal for how specific breakdowns in brain mechanisms of several different types, and in different parts of the brain, can interact together to cause a pervasive mental disorder. The article thus provides a way for many scientists who are not currently studying autism to better understand how autism may be related to concepts and data about normal brain and behavior with which they are familiar, and for specialists in autism to use concepts and data about normal brain and behavior to better understand symptoms of autism.

The article is organized as follows: Sections 2 and 3 review key data about autism. Sections 4-8 summarize key principles, mechanisms, and properties of three previously published brain models of normal behavior in a heuristic manner, and describe relevant quantitative properties of these models, with both normal and abnormal parameters, to clarify how autistic behaviors may be caused by prescribed breakdowns of normal mechanisms in several brain systems. These three models describe aspects of: (1) how the brain can learn to recognize objects and events, notably how it can determine the correct level of generalization with which to understand these objects and events in a context-appropriate manner (Carpenter and Grossberg, 1991, 1993; Grossberg, 1980, 1982b, 1984a, 1987, 1999b; Raizada and Grossberg, 2003); (2) how the brain can learn the emotional meanings of events through cognitive-emotional interactions, notably how such meanings can be learned through rewarding and punishing experiences, and how the brain can respond to emotional events with a properly calibrated intensity (Grossberg, 1972a, 1972b, 1980, 1982a, 1984a, 1984b, 2000b; Grossberg and

Gutowski, 1987; Grossberg and Levine, 1987; Grossberg and Schmajuk, 1987); and (3) how the brain can learn to adaptively time how long it focuses its attention upon motivationally important events, and when to respond to these events, in a context-appropriate manner (Bullock, Fiala, and Grossberg, 1994; Fiala, Grossberg, and Bullock, 1996; Grossberg and Merrill, 1992, 1996; Grossberg and Paine, 2000; Grossberg and Schmajuk, 1989). All three component models have been mathematically and computationally characterized elsewhere in order to explain behavioral and brain data about normal and, in some cases, abnormal behaviors. The principles and mechanisms that these models employ have thus been independently validated through their ability to explain data other than data about autism. The exposition herein will be kept as non-technical as possible in order to state the main ideas with a minimum of distracting detail.

Sections 7 and 8 also describe how these three models can be joined together in a larger brain system, the START model (Grossberg and Merrill, 1992, 1996), which has been used to quantitatively explain additional data about normal brain and behavior. Finally, these sections propose how changing parameters in the three component models of the START model can, when they ramify via interactions throughout the entire iSTART system, give rise to abnormal behavioral symptoms that strikingly resemble the behavioral symptoms of autistic individuals. This approach, for the first time, links brain mechanisms of normal behavior to proposed brain mechanisms of autism that are capable of giving rise to autistic behavioral symptoms, and makes detailed predictions about the types of brain mechanisms that may give rise to autistic behavioral symptoms.

Section 9 reviews other models of autism, states their connections with the iSTART model, comparatively analyses the explanatory advantages of the iSTART model over these previous proposals, and thereby shows how the iSTART model can provide a unifying framework for evaluating and further explicating many previously available models of autism.

METHODS AND MATERIALS

2. Key Features of Autism

Autism manifests during the first three years of life. The core features of autism (American Psychiatric Association, 1994) are qualitatively impaired socialization, impaired verbal and nonverbal communication, and restricted and repetitive patterns of behavior, interests, and activities. Children with autism are impaired in at least one or more of the following areas with onset before three years of age: social interaction; language as used in social communication; and symbolic or imaginative play.

Autism includes a spectrum, or broad range, of behavioral and cognitive manifestations. Mental retardation is common but between one fourth and one third of autistic individuals have IQs in the normal range or above (Fombonne, 1999). Approximately one third of autistic individuals develop epilepsy by adulthood (Tuchman and Rapin, 1997). Twenty to thirty percent of individuals with autism have apparent regression of skills after a period of apparent normal development, typically between one and three years of age (Tuchman and Rapin, 1997).

The term Autistic Spectrum Disorder (ASD) refers to this entire range of functions and findings within those individuals who have the key triad of impairments in social interaction, verbal and non-verbal communication, and restricted, repetitive, and stereotyped interests and activities. This term includes individuals who range from low-function, with profound impairments in many areas, to individuals with high-function. The Diagnostic and Statistical Manual of Mental Disorders (1994) defines autism as part of the group of Pervasive Developmental Disorders. Both Autistic Disorder and Pervasive Developmental Disorder-Not

Otherwise Specified/Atypical Autism (PDD-NOS) are generally considered part of autism. Asperger Syndrome shares the core deficits in social interactions and restricted/repetitive behaviors and interests, while relatively sparing formal language development. Asperger Syndrome individuals nevertheless have abnormal social communication skills. The term "autistic spectrum" was coined by Wing (1997) to include those diagnostic labels within the DSM IV and the International Classification of Diseases (1994) and other individuals with the key features of impairments in social interaction, impairments in verbal and non-verbal communication with a narrow/repetitive pattern of behaviors, and with impairments in imagination. Wing (1997, p. 1762) asserted that the extant diagnostic subdivisions were "arbitrary and … difficult to apply and unhelpful in clinical practice." Wing instead subdivided the spectrum by the nature of the social impairment and related co-morbidities: "aloof"; "passive"; "active but odd"; and "loners." While those subdivisions have not found broad usage, the concept of an autistic spectrum has found common if not universal acceptance. Many researchers and clinicians accept the concept (Tager-Flusberg *et al.*, 2001), although some caution that it implies stronger links between the disorders than current research may justify (Volkmar and Pauls, 2003). Also its common use may have led to excessive clinical labeling of some socially atypical children as autistic (Tager-Flusberg *et al.*, 2001). The goal of this paper is to understand how the range and common themes of the entire heterogeneous autistic spectrum, which will henceforth be referred to as "autism," can result from various combinations of disruptions in an established model of normal brain function. The model does this by proposing how breakdowns of multiple mechanisms in several brain areas may be involved in the generation of autistic behaviors. Differences in the degree of malfunction across these mechanisms may help to account for the differences in autistic populations. To accomplish this task we must first establish what clinical phenomena an adequate model of autism must explain.

An early manifestation of autism is a failure to develop basic imitation skills. Normal children usually show basic imitative behaviors by the end of the first year, and are imitating complex actions like wiping a table by fifteen months. This is not so of most autistic children (Receveur *et al*., 2005).

The communication deficits of autism often have onset before spoken language typically begins. Preverbal communication by way of gestures, sounds, and expressions are deficient. Many children with autism stay essentially non-verbal. Among those autistic individuals who develop language, other communication deficits follow and are not mere delays of the normal pattern of development, but are instead wide-ranging and complex disorders (Filipek *et al.*, 2000). Language pragmatics is often disturbed. Inappropriate and idiosyncratic word usages, such as inappropriate generalizations of meaning and odd analogies, are common. Often, individual words may be used with hyperspecificity and without ever being able to apply the word to a more general concept. Unusual intonation, echolalia, and pronoun reversal are common (Volkmar and Pauls, 2003).

Extreme unevenness in cognitive skills is a common feature of autism. Some autistic individuals have "islands" of normal or occasionally even superior ability and a few have narrow skill sets that are so superior to normal populations that they are referred to as "savants". These areas of higher ability often include mathematical and musical skills. Autistic individuals also have facilitated skills at detecting hidden, embedded figures (Jolliffe and Baron-Cohen, 1997; Ring *et al.,* 1999) and at copying "impossible figures" (Mottron *et al*., 1999). Thus, the autistic cognitive style is notable for its extreme concreteness, with autistic individuals doing relatively well on tasks that require rote memory but characteristically poorly on tasks that require higher-

order conceptual processes or abstraction (Filipek, *et al.* 2000). By eighteen months, an autistic child often lacks the normally emergent imaginary or symbolic play. Instead, an autistic child may tenaciously perseverate on specific features of a toy. Moreover in children with these tendencies, a favorite object must be exactly right (i.e., how it was when it was first noticed) and is often played with according to a very specific routine. It is as if each situation is learned as a complete specific whole and any variation from that standard invalidates any understanding that an autistic has of the situation and what to expect. When this "need for sameness" is violated by even minor variations in routines, behavioral decompensation in the form of strong emotional outbursts may result.

Extreme negative emotional reactivity is not only triggered by variations in routines; basic sensory stimuli, such as noise, smells, or light touch, can be emotional triggers for many as well. This contrasts dramatically with hyporesponsiveness to other, and in particular, social stimuli, such as one's name being called, facial reactions, or praise (Volkmar and Pauls, 2003; O'Neill and Jones, 1997).

Attentional differences are prevalent among patients with autism and include deficient "shared" or "joint" attention (Filipek *et al.,* 2000). Shared attention, which usually emerges during a normal child's first year of life, and refers to the ability to follow a significant other's gaze and thus to share attention in external objects with others, is characteristically deficient among those with autism. Reviews of home movies have documented attentional differences, with deficient attention in social but not in non-social areas, before six months of age in infants who were later to be identified as autistic (Maestro *et al.*, 2002). Autistic individuals also commonly experience difficulties with disengaging or shifting attention and with splitting attention between objects. Paradoxically, some autistic individuals clinically appear to have difficulty sustaining attention in certain contexts (Allen and Courchesne, 2001) and some studies have documented a significant frequency of symptoms severe enough to warrant labeling this behavior as comorbid Attention Deficit Disorder in some autistic spectrum individuals (Sturm, Fernell, and Gillberg, 2004; Yoshida and Uchityama, 2004; Goldstein and Schwebach, 2004).

Individuals with autism, especially low functioning individuals with autism, are also prone to repetitive stereotypic movements, such as rocking, hand flapping, and head banging. Other motor abnormalities variably include poor motor imitation abilities (Jones and Prior 1985), generalized clumsiness (Ghaziuddin and Butler, 1998), and gait abnormalities (Hallett *et al*. 1993; Vilensky *et al*., 1981). High functioning individuals with autism tend to have large handwriting even when controlled for educational level (Beversdorf *et al*., 2001). A retrospective review of videotapes (Teitelbaum *et al*., 1998) of infants later diagnosed to have autism suggested that movement abnormalities may be the earliest behavioral feature in autism, with abnormalities present by four to five months of age.

3. What Brain Abnormalities Cause Autism?

Various sorts of analysis have been employed to attempt to determine what brain abnormalities cause autism. These analyses have included cytoarchitectural studies, imaging of structural volumes, functional imaging studies, genetic and gene expression studies, lesion studies, and studies using behavioral paradigms known to be associated with particular brain structures. Systems implicated have included the cerebellar, limbic, and neocortical systems. The iSTART model clarifies how the various abnormalities that are summarized below all contribute to the creation of autistic behavioral symptoms via the interaction of different types of functional deficits across these brain areas.

Cytoarchitectual Studies. Postmortem cytoarchitecural studies have been limited by

small sample sizes but abnormalities of cerebellar structure have been frequently observed and consist of decreased Purkinje cell number and atypical Purkinje cell size and branching pattern without associated gliosis (Bailey *et al*., 1998; Baumann and Kemper, 1986; Fatemi *et al*., 2002; Kemper and Baumann, 1993; Ritvo *et al*., 1986). Cytoarchitectural abnormalities have also been documented, albeit with less consistency, in the amygdala, hippocampus, and ventral temporal cortex (Bachevalier and Loveland, 2003; Sweetten *et al*., 2002). The cytoarchitecture of cell columns in prefrontal and temporal cortex documents minicolumnar pathology in the brains of autistic individuals with more numerous and smaller cortical cell columns (Casanova *et al*., 2002).

Imaging Studies. Brain structures are not only affected at a microscopic level. Imaging studies have identified gross anatomical abnormalities in the frontal lobes, limbic systems, and cerebella of autistic individuals, although findings have been variable and inconsistent. An exhaustive review of this literature is beyond the scope of this article and can be found elsewhere (Herbert *et al.*, 2004; Sokol and Edwards-Brown, 2004); a representative selection follows.

Multiple MRI studies have documented hypoplasia of the cerebellar vermis in many individuals with autism and hyperplasia in a few (Courchesne *et al*., 1988, 1994; Hardan *et al*., 2001; Saitoh and Courchesne, 1998), while others have failed to replicate the findings (Manes *et al*., 1999; Piven *et al*., 1997). Amygdala enlargement has been found among individuals with autism (Howard *et al*., 2000; Sparks *et al*., 2002), as has reduced amygdala volumes in others (Aylward *et al.*, 1999; Pierce *et al*., 2001).

Brain volume growth is also abnormal: One retrospective study of autistic children study documented relatively small head circumferences at birth with a rapid increase between one and fourteen months of age (Courchesne, *et al.* 2003); by 2 to 4 years of age 37% of autistic children met criteria for developmental macrocephaly (Lainhart, *et al.* 1997); and an increased total brain volume was seen in studies of older autistic children up through age 8 to 12, but not in children over 12 (Aylward *et al.* 2002; Courchesne *et al.* 2001). In studies of high-functioning autistic boys, Herbert and others have used an MRI parcellation technique to document that this increased volume appears to primarily occur within the white matter (Herbert *et al.*, 2003) and seems to be specific to the radiate white matter (Herbert, 2004). Atypical neocortical asymmetries have also been documented at smaller units of analysis, with reversal of the normal language-related frontal cortex asymmetry (De Fosse *et al.*, 2004) and greater asymmetry in a variety of neocortical areas, with the greatest amount being noted in higher-order association cortices, and with a greater proportion of rightward asymmetric cortex than controls (Herbert *et al.,* 2005).

Functional imaging studies also have variable findings. High-functioning autistic adults were studied during explicit and implicit processing of emotional face expressions (Critchley *et al*., 2000) and were found, relative to controls, to not activate a cortical "face area" (the fusiform gyrus) when explicitly judging emotional facial expressions, or the left amygdala and left cerebellum during implicit processing of emotional facial expressions. Another study of face perception in adults with autism (Pierce *et al*., 2001) also found that autistic individuals had reduced or absent activation of the fusiform face gyrus, the inferior occipital cortex, the superior temporal cortex, and the amygdala, and instead activated a variety of individual specific cortical areas. Similar results were found using blood oxygen level-dependent signal changes (Hubl *et al.,* 2003) in which autistic individuals had lower signal strength in the fusiform face area, particularly during face processing, and higher signals in the medial occipital gyrus. They interpreted these face processing deficits not as a primary deficit in this area, but a result of

differences in how faces are processed – supporting models that hypothesize a preference for local over global modes of information processing in autism. The construct that decreased fusiform face activation in autistic individuals is a result of a different processing strategy than that of the normal population is further supported by a study in which autistic subjects processed familiar faces with a similar, albeit more limited, network as normal subjects, including fusiform face area and amygdala activation (Pierce *et al.,* 2004), and another study (Hadjikhani *et al.* 2004) that found no significant difference in fusiform face area activation between autistic spectrum disorder individuals and normal controls using a different task and measuring technique. These studies together suggest that the fusiform face area is functional in autistic individuals, but used differently than in the normal population in a task-dependent manner.

Abnormal fMRI results were also found in high functioning adults with autism while performing spatial working memory tasks (Luna *et al*., 2002), with specific reduction of activation in the dorsolateral prefrontal cortex relative to normal controls. High functioning adults with autism during a task involving making mentalistic inferences from viewing another's eyes (Baron-Cohen *et al*., 1999) exhibited relative reduction of activation in the amygdala and superior temporal cortex, and relative activation of fronto-temporal cortical areas. High-functioning autistic adults during an embedded figures task, which autistic individuals typically perform better than normal controls (Ring *et al*., 1999), showed less prefrontal activation and greater ventral occipitotemporal activation than normal controls. In a task involving visually-paced finger movements (Muller *et al*., 2001), there was reduced activation in areas activated in a control group and activation instead in differing areas in each individual with autism.

Other fMRI studies also demonstrated how autistic individuals use cortical resources in different ways than the normal population when solving various sorts of problems. During a sentence comprehension task, high-functioning autistic individuals activated Wernicke's area more, and Broca's area less, than the normal population. Moreover, various cortical areas were less synchronized in the autistic individuals, a finding that was felt by the authors to support "underconnectivity" as etiologic in autism (Just *et al*., 2004). A subsequent study (Koshino *et al*., 2005), however, was in conflict with such an interpretation. In that study, adults with high-functioning autism were again compared to a normal population, but this time during an n-back working memory task with letters, on which both groups performed comparably well. The group with autism showed greater activation of the right prefrontal and parietal areas, and more in the posterior regions, compared to the normal group's greater activation in the left parietal area. Significantly, there were no differences documented in the degree of synchronization of different regions between the groups, although the group with autism synchronized the prefrontal areas with the right parietal area, and the control group synchronized the prefrontal areas with the left parietal area.

Imaging studies of metabolic activity in the brains of those with autism also show various distortions and deficiencies. Magnetic resonance spectroscopy documented decreased cerebellar N-acetyl-aspartame (NAA, an indicator of neuronal function and maturity) in the cerebella of nine autistic children compared to five sibling controls in one study (Chugani *et al*., 1999). Twenty-seven autistic individuals compared to ten normal controls, with a similar protocol, were found to have decreased NAA in both the cerebellum and the hippocampal-amygdala region (Otsuka *et al*., 1999).

Genetic Studies. Genetic factors play a significant and complex role in autism (Rutter, 2000): Concordance rates were 60% for monozygotic twin pairs and 5% for dizygotic pairs; the relative increase risk for siblings was 30 to 100-fold; and a broader autistic phenotype consisting

of mixed patterns of cognitive and social deficits with repetitive stereotyped interest patterns was found in the families of those with autism. Analyses have suggested that multiple genes are involved: estimates range from between two to ten genes (Pickles et al., 2000) to more than fifteen genes, each with minor effects (Risch et al., 1999). These results support the hypothesis that autism may be caused by multiple functional deficits across several brain regions. Linkage analyses using molecular genetic techniques have not yet conclusively identified any individual genes, although many have been suggested (Lamb et al., 2000). Future research may clarify which genes are definitely involved, and where and when these genes are expressed, thereby helping to explain how different brain structures and their functional properties may be altered during autism.

Lesion Studies. Another approach to the identification of brain structures involved in autism has been through the use of both animal and human lesion studies. Early amygdala lesions have been studied in macaque monkeys (Machado and Bachevalier, 2003) and have resulted in alterations in the magnitude of fear responses beginning between three to five months after surgery: an increased fear response to peers, and a decreased fear response to novel objects. Macaque monkeys who had amygdala lesions during infancy also had cognitive and socio-emotional abnormalities as adults. Amygdalectomized animals explored the environment less, initiated fewer social interactions, and had some increased active social withdrawal. Humans with bilateral damage to the amygdala were found to judge people to look more trustworthy and more approachable than did normal subjects (Adolphs et al., 1998). Isolated bilateral hippocampal sclerosis of early childhood onset results in abnormal behavior and development (DeLong and Heinz, 1997) with failure of learning language and complex social and adaptive skills in general. Children who had surgical resection of cerebellar tumors (Levisohn et al., 2000) were documented to have impairments in executive function, including planning and sequencing, and in visual-spatial function, expressive language, verbal memory and modulation of affect. Vermal lesions were most associated with affective dysregulation, including blunted affect, and disinhibited and inappropriate behaviors.

Behavioral Studies. Established behavioral paradigms known to involve particular brain systems have also been used to demonstrate which brain systems are involved in autism. Several examples of this approach follow.

Abnormalities of performance occur during classical conditioning of the eye-blink response (Sears *et al.,* 1994). These abnormalities clarify the types of adaptive timing deficits that autistic individuals may have, and that may express themselves in a variety of behavioral contexts. The classically conditioned eye-blink response is a useful probe of these adaptive timing deficits because it is one of the best-studied examples of associative learning in vertebrates. In this paradigm, a sensory stimulus, such as a light or tone, is paired with a puff of air after some interval. The air puff unconditionally elicits an eye blink. Pairing leads to learning whereby the sensory stimulus can elicit a conditioned eye blink response at around the expected time of the air puff. It is known that the cerebellum helps to coordinate the timing and amplitude of the conditional eye-blink and that the cerebellar cortex is critically involved in its adaptive timing (Bullock *et al.*, 1994; Fiala *et al*., 1996; McCormick *et al*., 1982; McCormick and Thompson, 1984; Perrett *et al*., 1993). Compared to controls, autistic subjects learned the task more quickly, but performed short-latency, high-amplitude responses. In other words, the adaptive timing that is controlled by cerebellar cortex was absent. The conditioned eye-blink paradigm provides a way to test a cerebellar adaptive timing deficit that may express itself in a variety of more complex and socially important behaviors.

Difficulty in controlling the timing of responses was also found in a study that used a temporal reproduction paradigm (Szelag *et al.*, 2004). Autistic individuals were unable to modulate their responses to auditory and visual stimulus duration, and consistently reproduced the same response duration independent of stimulus duration. Again, the adaptive timing function that is controlled by the cerebellum was missing.

Oculomotor movement abnormalities have been variable. Depending on the study, they have been found to be consistent with cerebellar pathology (Takarae *et al.*, 2004), inconsistent with cerebellar dysfunction and more consistent with neocortical dysfunction (Minshew *et al*., 1999), or inconsistent with an intrinsic abnormality in any specific area and best explained by disturbances of functional connectivity within the visual pursuit system (Goldberg *et al.*, 2002; Nowinski *et al*., 2005). Some data from a study of autistic reflexive and volitional saccades are more consistent with prefrontal effects. No difference was found between the autistic and the control group on measures of saccade metrics, while there was an increase in response suppression errors on both anti-saccade and oculomotor delayed-response tasks (Minshew *f*., 1999).

Other studies document motor planning deficiencies in autistic individuals, which may be consistent with cerebellar dysfunction (Beversdorf *et al*., 2001; Rinehart *et al.*, 2001). One study showed that visual attention deficits, including slowed covert orienting of attention, are similar in both adult autistic individuals and other individuals with cerebellar damage (Townsend *et al.,* 1999). Functional MRI studies show cerebellar activation in tasks of visual selective attention and attention-shifting devoid of motor components (Allen *et al*., 1997; Le, 1998). Patients with cerebellar damage from strokes or from tumor are slow to orient attention in space, and approach normal performance only after a 800–1200 msec delay (Townsend *et al*., 1999). Similar slowed orienting of visual attention has been documented in children with autism and is correlated with the degree of cerebellar hypoplasia (Harris *et al.,* 1999).

Event-Related Potential Studies. Event-related potentials (ERPs) provide useful data about brain dynamics that correlate with higher cognitive functions and may be altered in individuals with autism. Such data may be clarified by the iSTART model because component models of iSTART predicted ERP properties of normal individuals that were later confirmed experimentally. The ERP P300 covaries with stimulus probability and task relevance. It was smaller in autistic children 8 to 19 years old than in age and gender matched controls for phonetic stimuli at left hemisphere recording sites and not at right hemisphere sites, while no differences were found for musical chord stimuli at any site (Dawson *et al*., 1988), consistent with the concept that adaptive timing for non-language auditory input (musical chord stimuli) develops normally, whereas processing of language-related auditory stimuli (phonetic stimuli) is reduced relative to normal levels.

In summary, autism is a heterogeneous condition with key behavioral features. These key features include impaired socialization, impaired verbal and nonverbal communication, and restricted and repetitive patterns of behavior, interests, and activities. Associated features include: impaired imitation skills; uneven cognitive skills with a concrete learning style and impairments in abstract thought and symbolic or imaginary play, and a range of cognitive ability from mentally retarded to normal or above average; diminished emotional reactivity to many (particularly socially salient) stimuli while paradoxically being prone to strong negative emotional reactivity to some lower-order stimuli or to variations in routine; attentional differences that include impaired joint attention and difficulties in flexibly disengaging and shifting attention while also occasionally being easily distractible; subtle motor anomalies; and

onset of subtle abnormalities during the first year of life and manifesting more dramatically during years two and three of life.

A complete model of autism needs to explain the genesis and progression of these key and associated features, be consistent with the behavioral heterogeneity within and across the disorder (inclusive of Autistic Disorder, Asperger's Syndrome, and Pervasive Developmental Disorder-Not Otherwise Specified), and should clarify how apparently many genotypes and neuroanatomic variations of the neocortical, limbic, and cerebellar systems are all associated with this similarly disordered developmental pathway (see Table 1).

Table 1

a) Some Key Symptoms of Autism that iSTART Helps to Explain

Social skill impairments
Verbal and nonverbal communication deficiencies
Restricted and repetitive patterns of behavior, interests, and activities
Imitation skill deficiencies
Uneven cognitive skills including a concrete learning style with impaired abstract thought
Variable levels of cognitive ability from mentally retarded to normal or above average
Diminished emotional reactivity to many (particularly socially salient) stimuli
Strong negative emotional reactivity to some lower-order stimuli or to variations in routine
Impaired joint attention and difficulties in flexibly disengaging and shifting attention
Deficient adaptive timing of motor behaviors
Subtle abnormalities during first year becoming more dramatic during years two and three
Abnormalities in neocortical, hippocampal, limbic, and cerebellar regions of the brain
Abnormalities in functional neuroimaging and event-related potential measures of brain processing

MODEL EXPLANATIONS

4. iSTART: A Neural Model of How Autistic Symptoms Can Arise

The data reviewed above summarize the case for characterizing autism as resulting from early onset dysfunction in parts of the neocortical, limbic and cerebellar systems. To mechanistically understand how these brain dysfunctions result in the behavioral and cognitive manifestations of autism requires, first, a neural model of the normal behavioral functions that break down during autism. The model used in this article is a system-wide synthesis of three neural models that have each previously been developed to explain different sorts of brain and behavioral data other than data about autism. After a brief overview, each component model will explained more fully and then integrated with the other models into a larger model system, called START (Spectrally

Timed Adaptive Resonance Theory). When parameters of START are suitably altered in the Imbalanced START, or iSTART, model, formal behavioral symptoms strikingly like those of autism are observed. The model also offers testable hypotheses about the brain mechanisms that cause the symptoms and how they break down.

The first model, called Adaptive Resonance Theory, or ART, proposes how the brain learns to recognize objects and events. This is accomplished through an interplay between bottom-up perceptually-driven inputs and learned top-down expectations. Bottom-up inputs attempt to match top-down expectations, and the top-down expectations can prime the brain to anticipate expected feature patterns. When a match occurs, the system locks into an attentive resonant state that drives the recognition learning process; hence the term *adaptive* resonance. ART predicts that *all conscious events are resonant events*. The degree of match that is required for resonance and sustained attention to occur is set by a *vigilance* parameter. Vigilance may be increased by predictive errors, and controls whether a particular learned representation will be concrete or abstract. Low vigilance allows the learning of broad abstract recognition categories; high vigilance forces the learning of specific concrete categories. If a match is inadequate, then the current input is processed as a novel stimulus. Attention is then rapidly reset so that memory can be searched for another, or new, representation of the event. The iSTART model proposes that individuals with autism have their vigilance fixed at such a high setting that their learned representations are very concrete, or hyperspecific; that is, *hypervigilance leads to hyperspecific learning*. This property leads to a multitude of problems with learning, cognition, and attention due to the manner in which top-down matching, attention, learning, attention focusing and reset, and memory search are organized. The ART model clarifies thalamo-cortical-hippocampal interactions, among others in the brain.

The second model, called the CogEM (or Cognitive-Emotional-Motor) model, extends ART to the learning of cognitive-emotional associations, notably associations that link external objects and events in the world to internal feelings and emotions that give these objects and events value. These emotions also activate the motivational pathways that energize actions aimed at acquiring or manipulating objects or events to satisfy them. Resonance can also occur within CogEM circuits. Here the resonance is between sensory/cognitive representations of what is possible and emotional representations of what is valued. The resonance tends to focus attention selectively upon objects and events that promise to satisfy emotional needs. The emotional representations are proposed to be organized into opponent affects, such as fear vs. relief. These opponent affective circuits are energized by internal sources of arousal. Under normal circumstances, arousal is set at an intermediate level; that is, a Golden Mean is typically maintained at, or near, an intermediate arousal level, at least most of the time while we are awake. Either underarousal or overarousal can cause abnormal emotional reactions and, with them, abnormal cognitive-emotional learning. In particular, there is *an Inverted-U* in emotional reactivity in these opponent circuits. If the emotional center is overaroused, then the threshold to activate an emotion is abnormally low, but the intensity of emotion is abnormally small. In contrast, if the emotional circuits are underaroused, then the threshold for activating an emotion is abnormally high but, when this threshold is exceeded, the emotional response can be hyperreactive. The iSTART model proposes that various individuals with autism experience underaroused emotional depression, which helps to explain their emotional outbursts as well as their baseline of reduced emotional expression. The CogEM model clarifies thalamo-cortico-amygdala interactions, among others in the brain.

The third model, called the Spectral Timing model, clarifies how the brain adaptively times responses in order to acquire rewards and other goal objects. Such adaptive timing is essential for all terrestrial animals, since rewards and other goals are often delayed in time relative to actions that are aimed at acquiring them. The brain needs to be dynamically buffered, or protected against, reacting within the time interval before a delayed reward can be received as a predictive failure. The Spectral Timing model accomplishes this by predicting how the brain distinguishes *expected non-occurrences* of rewards, which should not be allowed to interfere with acquiring the delayed goal, from *unexpected non-occurrences* of rewards, which can trigger the usual consequences of predictive failure, including reset of working memory, attention shifts, emotional rebounds, and the release of exploratory behaviors. The Spectral Timing model proposes how various individuals with autism experience failures of adaptive timing, thus leading to the premature release of behaviors in a context-inappropriate manner that can prevent these behaviors from being rewarded. The Spectral Timing model clarifies thalamo-cortico-hippocampal-cerebellar-basal ganglia interactions, among others in the brain.

The next section reviews these models of how the brain gives rise to normal behaviors in more detail. This review summarizes key properties and mechanisms of each model and illustrative behavioral and brain data about normal individuals that each model has successfully explained and predicted. The fact that these models have explained and predicted more behavioral and brain data than competing models provides some confidence in their underlying brain mechanisms. It is then shown how these models can generate formal symptoms that resemble autistic behaviors when their mechanisms are imbalanced and/or lesioned in prescribed ways as part of the iSTART model. This exposition hereby links autistic symptoms to mechanisms and data that have no obvious connection to autism, and thereby opens up a wide range of possible new experiments to evaluate autistic behaviors and to further test and develop the model. Such an approach also should make it easier for scientists who are studying several types of normal behavioral and brain properties to link their work to studies of autism.

5. Adaptive Resonance Theory model

5.1. Perceptual and Cognitive Learning, Expectation, Attention, and Fantasy

First let us consider the Adaptive Resonance theory, or ART, model. ART proposes an answer to the "stability-plasticity dilemma"; namely, how the brain can learn quickly throughout life without being forced to unselectively forget previously learned memories just as quickly (Grossberg, 1980, 1999b). This problem has also been called the catastrophic forgetting problem (Carpenter, 2001; French, 1999; Page, 2000). ART proposes how normal learning and memory may be stabilized through the use of learned top-down expectations (Figure 1a). In other words, we are "intentional" beings so that we can learn quickly without suffering catastrophic forgetting. Top-down expectations have been predicted to operate at multiple cortical and thalamic levels, including top-down expectations from higher cortical areas such as from prefrontal to inferotemporal cortex, and also at lower cortical areas such as from striate visual cortex to the lateral geniculate nucleus (Gove, Grossberg, and Mingolla, 1995; Grossberg, 1999b, 2003a; Raizada and Grossberg, 2003).

Top-down expectations learn prototypes that are capable of focusing attention (Figure 1b) upon the combinations of features that comprise conscious perceptual experiences. When top-down expectations are active in a priming situation in the absence of bottom-up information, they can modulate or sensitize their target cells to respond more effectively to future bottom-up information that matches the prototype. Such expectations cannot, however, fully activate these target cells under most circumstances. When bottom-up inputs do occur, an active top-down

expectation selects the cells whose input features are consistent with the active prototype, and suppresses those that are not, thereby generating an attentional focus on the combinations of features that may be expected in that situation. This matching and attentional process can synchronize and amplify the activities of selected cells, leading to a context-sensitive state of "resonance". Thus, attentionally relevant stimuli are selected for learning, while irrelevant stimuli are suppressed and hence prevented from destabilizing existing representations. Such a matching process has been mathematically proved to be necessary to stabilize the memory of learned representations in response to a complex input environment (e.g., Carpenter and Grossberg, 1991).

The solution that ART proposes to the stability-plasticity dilemma is to allow neural representations to be rapidly modified only by those incoming stimuli with which they form a sufficiently close match. If the match is close enough, then learning occurs. Precisely because the match is sufficiently close, this learning causes fine-tuning of the existing representation, rather than a radical overwriting. Matching gets started by initially endowing the top-down matching pathways with broadly distributed adaptive weights. Learning prunes these weights and makes them more selective. If the active neural representation does not match with the incoming stimulus, then its neural activity will be inhibited and hence unable to cause plastic changes.

In order to realize these matching properties, top-down expectations and attention were predicted to be controlled by top-down on-center off-surround networks (Figure 1c). A balance between top-down excitation and inhibition in the on-center of this network enables attention to provide excitatory modulation of on-center cell responses to bottom-up inputs, even while cells are strongly inhibited in the off-surround. Recent psychophysical and neurophysiological data have supported many ART predictions about the link between learning, matching, competition, and attention.

The prediction that top-down attention has an on-center off-surround characteristic has received psychological and neurobiological empirical confirmation in the visual system (Bullier, Hupé, James, and Girard, 1996; Caputo and Guerra, 1998; Downing, 1988; Mounts, 2000; Reynolds, Chelazzi, and Desimone, 1999; Smith, Singh, and Greenlee, 2000; Somers, Dale, Seiffert, and Tootell, 1999; Sillito, Jones, Gerstein, and West, 1994; Steinman, Steinman, and Lehmkuhle, 1995; Vanduffell, Tootell, and Orban, 2000). Feedback from auditory cortex to the medial geniculate nucleus (MGN) and the inferior colliculus (IC) also has an on-center off-surround form (Zhang *et al.*, 1997), as does feedback within the rodent barrel system (Temereanca and Simons, 2001). Based on such data, the ART attentional prediction has recently been described using the phrase "biased competition" (Desimone, 1998; Kastner and Ungerleider, 2001), in which attention biases the competitive influences within the network.

The attentional feedback pathway has recently been predicted to occur in identified cells and cell layers within the neocortex, notably within the visual cortex. This refinement of ART is called the LAMINART model (Grossberg, 1999a, 2003a; Raizada and Grossberg, 2003). The LAMINART model predicts that top-down signals from layer 6 of a higher cortical area reach layer 6 of a lower cortical area, where they are relayed to layer 4 of that area via modulatory on-center off-surround interactions.

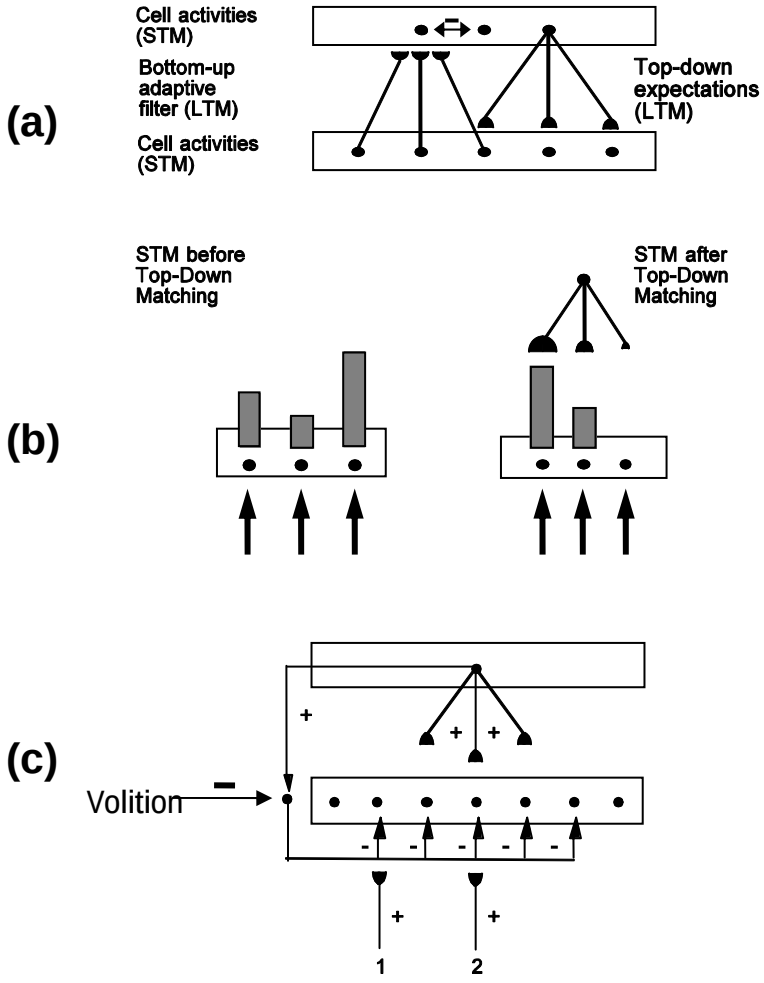


Figure 1. (a) Patterns of activation, or short-term memory (STM), across feature-selective cells at a lower processing level send signals via bottom-up pathways to a higher processing level. Cells at the higher level respond selectively to prescribed combinations of features at the lower level. For example, such cells may represent recognition categories, as in inferotemporal cortex. The selective activation of category cells is achieved by multiplying the bottom-up signals with adaptive weights, or learned long-term memory (LTM) traces at the ends of the bottom-up pathways, before these learning-gated signals activate target category cells. These category cells compete among themselves to select a small number of winning cells. The combination of bottom-up adaptive filtering and competition are the basic ones for defining a self-organizing map; see Section 14. The active category cells, in turn, activate top-down pathways that read-out learned expectations via their own LTM traces. These top-down expectations are matched against the STM pattern that is active at the lower featural level. (b) This matching process confirms, synchronizes, and amplifies STM activities of features that are supported by large LTM traces in an active top-down expectation, and suppresses STM activities of features that do not get top-down support. The size of the hemidisks at the end of the top-down pathways represents the strength of the learned LTM trace that is stored in that pathway. (c) The ART Matching Rule may be realized by a modulatory top-down on-center off-surround network. In particular, bottom-up inputs, such as in pathways 1 and 2, can activate their feature-selective cells when no top-down expectation is active. When a top-down expectation is active whose prototype (the learned on-center with excitatory pathways) does not include the feature activated by pathway 1, then the top-down off-surround cancels the bottom-up input, thereby suppressing activation of that feature. Since the feature that is activated by pathway 2 is included in the top-down prototype, the top-down excitation and inhibition approximately cancel (typically, with a small positive priming bias), so that activation of the corresponding feature-selective cell is preserved, synchronized, and even amplified. [Reprinted with permission from Grossberg (1999b).]

The ART prediction that bottom-up sensory activity is *enhanced* when matched by top-down signals is in accord with an extensive neurophysiological literature showing the facilitatory effect of attentional feedback (e.g., Luck *et al.*, 1997; Roelfsema *et al*., 1998) but not with models in which matches with top-down feedback cause suppression (e.g., Mumford, 1992; Rao and Ballard, 1999). The ART prediction raises an additional question: Is there evidence that top-down feedback controls plasticity in the area to which it is directed?

Recent data support the ART prediction that top-down feedback regulates plasticity. Psychophysically, the role of attention in controlling adult plasticity and perceptual learning was demonstrated by Ahissar and Hochstein (1993). Neurophysiological evidence of Gao and Suga (1998) showed that acoustic stimuli caused plastic changes in the inferior colliculus of bats only when the IC received top-down feedback from auditory cortex. This plasticity is enhanced when the auditory stimuli were made behaviorally relevant, in accord with the ART proposal that top-down feedback allows attended—that is, relevant—stimuli to be learned, while suppressing unattended irrelevant ones. Evidence that cortical feedback also controls thalamic plasticity in the somatosensory system has been found by Nicolelis and colleagues (Krupa *et al*., 1999) and by Parker and Dostrobsky (1999). See Kaas (1999) for a review.

Another predicted role of these feedback connections is to synchronize the firing patterns of higher and lower cortical areas. Given that "cells that fire together wire together", synchronous firing would be expected to increase the ability of the mutually excitatory resonant activity caused by ART matching to facilitate synaptic plasticity and learning. It has elsewhere been shown that variants of the ART and LAMINART models are capable of rapidly synchronizing their activation patterns during both perceptual grouping and attentional focusing (Grossberg and Somers, 1991; Grossberg and Grunewald, 1997; Yazdanbakhsh and Grossberg, 2004). Recent discussions of top-down cortical feedback, synchrony, and how they support ART predictions are given by Engel, Fries, and Singer (2001), Fries, Reynolds, Rorie, and Desimone (2001) and Pollen (2001).

The ART model predicts how the brain has exploited the modulatory property of expectations and attention to enable fantasy, imagery, and planning activities to occur (Grossberg, 2000a). In particular, phasic *volitional signals* can shift the balance between excitation and inhibition by increasing the gain of excitation and inhibition when a top-down expectation is active (Figure 1c). Such a volitionally-mediated shift enables top-down expectations, in the absence of supportive bottom-up inputs, to cause conscious experiences of imagery and inner speech, and thereby to enable fantasy and planning activities to occur. If, however, these volitional signals become tonically hyperactive during a mental disorder, top-down expectations are predicted to give rise to conscious experiences, such as hallucinations, in the absence of bottom-up inputs and volition. The ability of top-down expectations to activate internal representations that support imagery, fantasy, and planning activities raises the issue of how these expectations are themselves controlled. Below it is suggested how interactions between cognitive-emotional mechanisms from CogEM and of cognitive and perceptual mechanisms of ART help to clarify how this happens.

Before turning to these interactions, it is worthwhile to mention a basic property of ART which the iSTART model identifies as being fundamental in autism. This property concerns the manner in which brain learning controls the level of abstractness of learned prototypes. This issue is of particular importance in light of the concreteness and hyperspecificity of autistic cognitive processing.

5.2. How is the Generality of Knowledge Controlled? Exemplars, Prototypes, and Vigilance

What information is bound together into object or event representations? Some evidence suggests that exemplars, or individual experiences, can be learned and remembered, like those of familiar faces (Medin, Altom, and Murphy, 1984; Medin and Shaffer, 1978; Medin and Smith, 1981). However, this cannot be the final answer to this question, since storing every exemplar in memory can lead to a combinatorial explosion of memory storage, unwieldy memory retrieval, and an inability to learn general or abstract properties of the world. Other scientists have proposed that we learn prototypes (Posner and Keele, 1970; Smith and Minda, 1998, 2000; Smith, Murray, and Minda, 1997) that represent more general properties of the world, such as the fact that everyone has a face. But then how do we learn specific episodic memories, and how is the appropriate level of generalization and abstraction determined?

ART provides an answer to this question that overcomes these problems and clarifies how the inferotemporal cortex, interacting with prefrontal cortex and the hippocampal system, learns to recognize and classify objects and events. In particular, one class of thirty human cognitive experiments (the so-called 5–4 category structure) has been used to test conflicting views in the prototype-exemplar debate (Medin, Altom, and Murphy, 1984; Nosofsky, 2000; Nosofsky, Kruschke, and McKinley, 1992; Nosofsky and Zaki, 2002; Smith and Minda, 2000), but exemplar and prototype models have not explained how these recognition categories are learned. Neurophysiology labs have also collected data about monkey cell responses from inferotemporal cortex during recognition tasks, showing that both specific and general information can be represented in this brain region (Desimone, 1991; Desimone and Ungerleider, 1989; Gochin, Miller, Gross, and Gerstein, 1991; Harries and Perrett, 1991; Mishkin, 1978, 1982; Mishkin and Appenzeller, 1987; Perrett, Mistlin, and Chitty, 1987; Schwartz, Desimone, Albright, and Gross, 1983).

An ART model has been developed that quantitatively simulates the pattern of human data from the 5–4 category structure experiments and clarifies neurophysiological data about how monkeys learn to categorize both exemplars and prototypes (Carpenter and Grossberg, 1991, 1993; Ersoy, Carpenter, and Grossberg, 2002; Grossberg, 1980, 1999b; Grossberg, Carpenter, and Ersoy, 2005). In particular, Carpenter, Ersoy, and Grossberg have shown that, when popular exemplar models are interpreted as processing models that carry out their operations in real time, they implicitly invoke categories which bind together the exemplars that belong in each category, as well as both bottom-up and top-down learned interactions among each category and its exemplars. Thus, even exemplar models, when interpreted as real-time processing models, implicitly embody many of the properties that ART models make explicit in order to show how categories and their top-down expectations are learned through time without experiencing catastrophic forgetting.

To overcome the problems of traditional exemplar and prototype models, ART learns prototypes that consist of the *critical feature patterns*, or combinations of relevant features to which an individual pays attention. These are not the traditional prototypes of prototype models, which are typically selected by experimentalists to define their experimental inputs. Rather, these critical feature patterns are incrementally learned by an ART model through time from sequences of experienced exemplars. ART mechanisms explain how an individual can autonomously discover which features are worthy of attention in any given situation. Such discovery can occur with or without external supervision. The generality of learned prototypes is determined by the network's *vigilance* parameter, which is controlled by environmental feedback or internal volition. Low vigilance permits learning of general categories with abstract prototypes. High

vigilance forces memory searches to occur for a new category when even small mismatches exist between an exemplar and the category that it activates. Given high enough vigilance, a category prototype may encode a single exemplar. ART proposes how, in a normal brain, vigilance can track the demands of a particular environment, creating specific or general categories as needed to solve environmental problems; see below. In the ART simulation of 5-4 category learning, a small number of abstract and specific critical feature patterns jointly form a distributed code for each learned category, thereby clarifying how a small number of cells in inferotemporal cortex can represent an object, and supporting the Nosofsky view that rules-plus-exceptions are learned (Nosofsky, 1984, 1987; Nosofsky, Kruschke, and McKinley, 1992; Palmeri, Nosofsky, and McKinley, 1994).

Spitzer, Desimone, and Moran (1988) have reported neurophysiological data that are consistent with the existence of vigilance control in the inferotemporal cortex, indeed with the ART prediction that links vigilance control, top-down matching, and attention. These authors write that "In the difficult condition the animals adopted a stricter internal criterion for discriminating matching from non-matching stimuli. The animal's internal representations of the stimuli were better separated…increased effort appeared to cause enhancement of the responses and sharpened selectivity for attended stimuli."

In a network whose vigilance is fixed through time at an abnormally high level, the system would be literally "hypervigilant," and environmental events would be classified with extreme concreteness and hyperspecificity, with learned categories coding highly specific, exemplar-like information. The iSTART model predicts that many individuals with autism have their vigilance fixed at a high level.

There are other mental disorders for which ART has proposed explanations in terms of defective vigilance control. Indeed, the ART model has earlier been used to explain data about the type of abnormal learning and memory that occur during medial temporal amnesia. Carpenter and Grossberg (1993; see also Grossberg and Merrill, 1996) predicted how a lesion of the ART orienting system, which is interpreted to model aspects of hippocampal dynamics, eliminates vigilance control; that is, the lesioned model behaves as if it has a very low vigilance. The model as a whole then exhibits many symptoms of medial temporal amnesia.

Recent support for this prediction has been forthcoming through the modeling work of Nosofky and his colleagues. Knowlton and Squire (1993) reported dissociations between categorization and recognition in amnesic individuals and used these data to argue for multiple memory systems to mediate these tasks. However, Nosofsky and Zaki (1998) and Zaki *et al* (2003) have shown that they can quantitatively fit the Knowlton and Squire (1993) and their own data using an exemplar model in which they choose a low value of their sensitivity parameter. Their low sensitivity parameter plays a role like the low vigilance parameter in ART. It should be noted that, when an exemplar model is interpreted as a real-time dynamical processing model, its hypotheses look very much like those of an ART model. These parallel approaches may thus become even more closely linked through future research. Indeed, Nosofsky and Johnson (2000, p.375) have argued that many multiple-system accounts can be replaced by a single system model when "similarity relations among exemplars change systematically because of selective attention to dimensions and because of changes in the level of sensitivity relating judged similarity to distance in psychological space. Adaptive learning principles may help explain the systematic influence of the selective attention process and of modulation in sensitivity settings on judged similarity." ART provides a dynamical account of how subjects can incrementally

learn to selectively pay attention to stimulus dimensions and of how they may alter their vigilance, or sensitivity, in a context-sensitive way.

In summary, just as abnormally low vigilance may help to explain properties of data about medial temporal amnesia, abnormally high vigilance may help to explain data about autism. The normal control of vigilance is thus a topic that warrants a great deal more neurobiological investigation.

5.3. How are Learning, Attention, Memory Search, Hypervigilance, and Hyperspecificity Related?

Given that vigilance control can enable a learning individual to learn either abstract and general, or concrete and specific, information as a particular learning environment demands, it is important to understand how vigilance control is realized under normal circumstances. Vigilance control is part of the process whereby top-down expectations match incoming bottom-up information, and determines whether a match is deemed good enough to trigger new learning. Figure 2 summarizes the ART proposal of how an *attentional system* and an *orienting system* normally work together to discover and learn effective recognition categories without experiencing catastrophic forgetting. As noted above, ART locates the attentional system in thalamocortical structures such as the temporal and prefrontal cortices. ART proposes that the orienting system includes the hippocampal system, which has long been known to be involved in mismatch processing, including the processing of novel events (e.g., Otto and Eichenbaum, 1992; Vinogradova, 2001).

Vigilance comes into the story during the matching process that takes place between an exemplar input and the top-down expectation that is read out by an active recognition category. A sufficiently bad mismatch is predicted to reset an active category, and to thereby initiate a memory search, or hypothesis testing, cycle that can lead to the selection of a new category. Such a mismatch can occur, say, because the exemplar input represents an unfamiliar type of experience. The new category is selected by a mismatch-driven memory search, or bout of hypothesis testing: A mismatch within the attentional system activates the complementary orienting system. Novelty-activated nonspecific arousal signals from the orienting system rapidly reset the recognition category within the attentional system that has been reading out the poorly matching top-down expectation (Figures 2b and 2c). The cause of the mismatch is hereby removed, thereby freeing the system to activate a different recognition category (Figure 2d). And so the cycle goes. If no matching recognition category exists, say because the bottom-up input represents a truly novel experience, then the search process automatically activates an as yet uncommitted population of cells, with which to learn a new recognition category to represent the novel information.

This hypothesis-testing cycle predicted that certain brain operations occur in a particular sequence of events. This prediction was tested by measuring sequences of event-related potentials that are recorded through scalp electrodes when humans perform recognition tasks (Banquet and Grossberg, 1987). In particular, correlated sequences of P120, N200, and P300 ERPs have been recorded which have the properties predicted by ART of mismatch, arousal, and reset operations, respectively. All breakdowns of ART search that are predicted below to occur individuals with autism should be reflected in these ERP measures.

In addition to this ERP support for the ART hypothesis-testing cycle, Miller, Li, and Desimone (1991) have reported neurophysiological data showing that there is an “active

matching process that was reset between trials" in monkey inferotemporal cortical cells that are involved in recognizing visual objects.

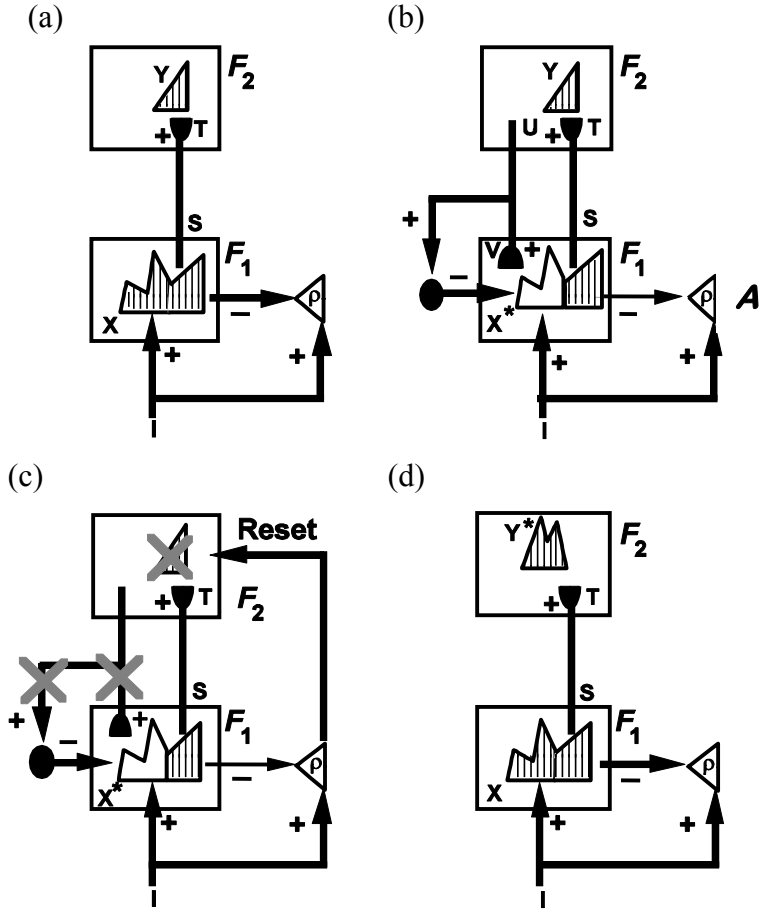


Figure 2. Search for and learning of a recognition code within Adaptive Resonance Theory: (a) The input pattern I is instated across the feature detectors at level F_1 as a short term memory (STM) activity pattern X. Input I also nonspecifically activates the orienting system A. STM pattern X is represented by the gray pattern across F_1. Pattern X both inhibits A and generates the output pattern S. Pattern S is multiplied by long term memory (LTM) traces, or learned adaptive weights. These LTM-gated signals are added at F_2 nodes to form the input pattern T, which activates the contrast-enhanced STM pattern Y across the recognition categories coded at level F_2. (b) Pattern Y generates the top-down output pattern U which is multiplied by top-down LTM traces and added at F_1 nodes to form the prototype pattern V that encodes the learned expectation of the active F_2 nodes. If V mismatches I at F_1, then a new STM activity pattern X^* is generated at F_1. X^* is represented by the gray pattern. It includes the features of I that are confirmed by the top-down expectation V. Mismatched features are inhibited. The inactivated nodes corresponding to unconfirmed features of X are unhatched. The reduction in total STM activity which occurs when X is transformed into X^* causes a decrease in the total inhibition from F_1 to A. (c) If inhibition decreases sufficiently, A releases a nonspecific arousal wave to F_2, which resets the categorical STM pattern Y at F_2. (d) After Y is inhibited, its top-down prototype signal is eliminated, and activity pattern X can be reinstated at F_1. Enduring traces of the prior reset lead X to activate a different STM pattern Y^* at F_2. If the top-down prototype due to Y^* also mismatches I at F_1, then the search for an F_2 code continues until a more appropriate F_2 representation is selected. Then an attentive resonance develops and learning of the attended data is initiated. [Adapted with permission from Grossberg and Merrill (1996).]

Vigilance ρ is computed within the ART orienting system *A* (ρ in Figure 2; see triangular symbol that represents *A*). Here, bottom-up excitation in an exemplar input pattern *I* activates feature-selective cells in the attentional system (labeled F_1 in Figure 2a), while also sending convergent activation to a population of cells in the orienting system. This distributed pattern of active cells in F_1 is denoted by *X* in Figure 2a. If more feature-selective cells are activated in the attentional system, then the total excitation of the orienting system increases too. What, then, keeps the orienting system from firing in response to all input patterns? When only the bottom-up exemplar input activates F_1, then all the active feature-selective cells can send convergent inhibition to the orienting system. The orienting system then remains quiet because its total excitatory input is balanced by its total inhibitory input, no matter how many cells may be activated by the input pattern.

When a top-down expectation also acts on F_1 , then only the "matched features" can remain active there, due to the ART Matching Rule (Figures 1c and 2b). That is, the activity pattern *X* caused solely by the bottom-up inputs *I* across the feature-selective cells (Figure 2a) is transformed into the pattern *X** of activation across the matched features (Figure 2b). Since fewer cells are active in *X** than in *X*, the total inhibition to the orienting system is reduced. This reduction in inhibition becomes greater as the mismatch between the bottom-up input pattern and the prototype of the top-down expectation becomes greater. If the mismatch becomes too great to satisfy the vigilance criterion, then a reset, or "novelty", wave is activated (Figure 2c).

How does the novelty wave know which cells in the category level F_2 need to be reset? The orienting system can compute that a mismatch has taken place in the attentional system, but not which categories in the attentional system read out top-down expectations that led to the mismatch. Thus, the reset wave takes the form of a burst of *nonspecific arousal* that equally arouses all of the recognition categories in F_2. The active categories F_2 respond to this arousal burst by being selectively inhibited by it, as in Figure 2c. Reset is accomplished by the same sort of interaction between arousal, competition, and habituative dynamics that is described more fully in the summary of cognitive-emotional interactions in Section 6 and Appendix A. (See Grossberg (1980) and Carpenter and Grossberg (1990, 1991) for descriptions of how selective reset happens in an ART cognitive learning network.) The mismatched top-down prototype is hereby inactivated and a search begins for another, better-matching, recognition category, as in Figure 2d.

More specifically, vigilance ρ weighs how similar an input exemplar *I* must be to a top-down prototype *V* in order for resonance to occur. Vigilance ρ is the relative gain of excitation to inhibition in the orienting system. For example, resonance occurs if ρ times the total bottom-up excitation from input *I* to the orienting system *A* is less than the total inhibition from *X** to the orienting system. Then the orienting system is inhibited, and resonance between levels F_1 and F_2 has a chance to develop. If, however, this inequality is reversed, then the orienting system can be activated, leading to a nonspecific novelty wave (e.g., an N200 ERP) that can reset the presently active category in F_2 and initiate search for a better-matching category.

ART proposes how vigilance gain criterion can be adjusted up and down within the orienting system to learn more specific or general information, respectively, in response to predictive failures within each environment. For example, suppose that a predictive failure causes vigilance to increase just enough to trigger the release of a novelty wave. Recall that low vigilance allows general categories to form, whereas high vigilance forces specific categories to form. Thus, every increase in vigilance implies that a more specific category will be learned. By bumping vigilance up to the smallest value that can correct a predictive error, the most general

categories can be learned that can eliminate predictive errors. This patented concept is called Match Tracking because vigilance tracks the match value. In this way, ART clarifies how the brain can try to learn categories that are sensitive to the statistical structure of any given situation, much as some cells in inferotemporal cortex learn to code highly specific information (e.g., a particular view of a familiar face), whereas other cells learn to code more general information about the environment. In general, combinations of both specific and general categories will be learned, so that both exemplars and more abstract information can be recognized.

If for any reason the vigilance gain gets "stuck" at a high level, then concreteness, or hyperspecific learning, will ensue. Persistently high vigilance can cause the learning of concrete and hyperspecific category prototypes both in the bottom-up filtering pathways that select a recognition category and in the top-down expectation pathways that focus attention upon expected events. Attentional deficits will ensue when vigilance remains abnormally high because hyperspecific prototypes will mismatch and reset attention in response to even small environmental variations.

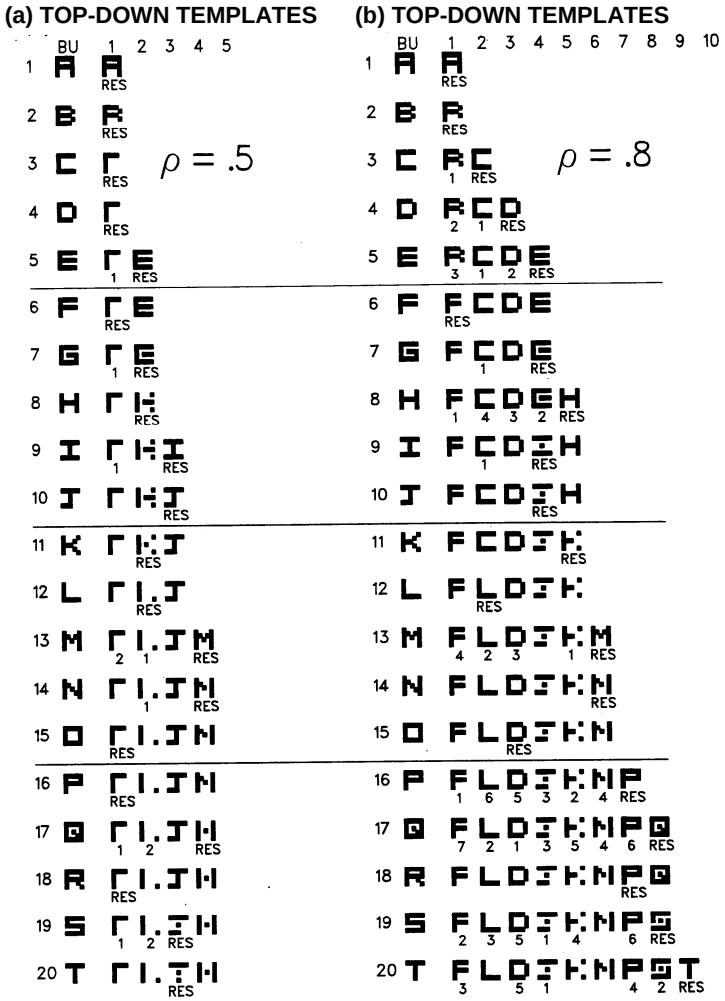


Figure 3. Alphabet learning by ART 1: Different vigilance levels (.5 and .8) cause different numbers of letter categories to be learned with different critical feature patterns. The critical feature patterns resemble the input exemplars more as vigilance is chosen higher. [Reprinted with permission from Carpenter and Grossberg (1987).]

Figure 3 summarizes a computer simulation from Carpenter and Grossberg (1987) that illustrates how ART can learn more general categories when vigilance is lower, and more specific categories when vigilance is higher, with the limit of maximal vigilance learning categories that include a single input exemplar. To simplify the example, there is no preprocessing of input letters as the alphabet is presented in order. The rows designate the learning trial (1, 2,…). The columns designate the inputs (BU, which abbreviates Bottom Up), the category number (1, 2,…), and the prototype, or template, that is learned on each trial. The symbol RES indicates that a resonance occurs between the input exemplar and the corresponding prototype. The numbers under the prototypes indicate the order in which the categories were searched before resonance occurred. In Figure 3a, vigilance is set at .5, and 4 categories are learned after 20 letters are presented. The prototypes are indicated in the last row of the figure. Their critical feature patterns are quite abstract. In Figure 3b, vigilance is set at .8, and 9 categories are learned after 20 letters are presented, with prototypes like look more like the input exemplars themselves. With vigilance set at 1, there would be 20 categories, each with an exemplar prototype.

5.4. Reconciling distributed and symbolic representations using resonance: Symbol grounding

If the top-down expectation is close enough to the bottom-up input pattern, then the pattern X^* of attended features reactivates the category Y which, in turn, reactivates X^*. The network hereby locks into a resonant state through a positive feedback loop that dynamically links, or binds, the attended features across X^* with their category, or symbol, Y.

The individual features at F_1 have no meaning on their own, just like the pixels in a picture are meaningless one-by-one. The category, or symbol, in F_2 is sensitive to the global patterning of these features, but it cannot represent the "contents" of the experience, including their conscious qualia, due to the very fact that a category is a compressed, or "symbolic" representation. Harnad (1990) has called this the "symbol grounding problem". It has often been erroneously claimed by practitioners of Artificial Intelligence that a single system can either process distributed features or symbolic representations, but not both. This is not true in the brain. Nor is it true in ART, which proposed a framework for solving the symbol grounding problem before Harnad described this problem, namely: The resonance between distributed feature patterns and recognition categories converts the *pattern* of attended features into a coherent context-sensitive state that is linked to its category through feedback. It is this coherent state, that joins together distributed features and symbolic categories into a unified bound state, which can enter consciousness. ART predicts that *all conscious states are resonant states*. In particular, such a resonance binds spatially distributed features into either a synchronous equilibrium or oscillation, until it is dynamically reset. Such synchronous states have recently attracted much interest after being reported in neurophysiological experiments; see Section 5.1. Synchronous states as a substrate of conscious experience were predicted in the 1970's in the articles which introduced ART. Grossberg (1999b) reviews this historical background and summarizes examples from the visual and auditory modalities that support the prediction that all conscious states are resonant states. The prediction that hyperspectific recognition categories may be learned by some individuals with autism implies that synchronous resonances may be reset more frequently in these individuals.

6. Cognitive-Emotional-Motor Model

6.1. Three Types of Representations and Learning

The second model is the CogEM model, which joins together Cognitive, Emotional, and Motor

processes (Grossberg, 1982a, 1984a, 2000b); see Figures 4 and 5. The CogEM model proposes how emotional centers of the brain, such as the amygdala, interact with sensory and prefrontal cortices — notably ventral, or orbital, prefrontal cortex — to generate affective states, attend to motivationally salient sensory events, and elicit motivated behaviors. Activating the feedback loop between cognitive and emotional centers is predicted to generate a cognitive-emotional resonance that can support conscious awareness of events happening in the world and how we feel about them. Recent experimental data provide increasing support for the predicted role of interactions between amygdala and orbitofrontal cortex in the control of response selection and predicted outcomes based on value acquired through previously rewarded behaviors (Baxter *et al*., 2000; Schoenbaum *et al*., 2003).

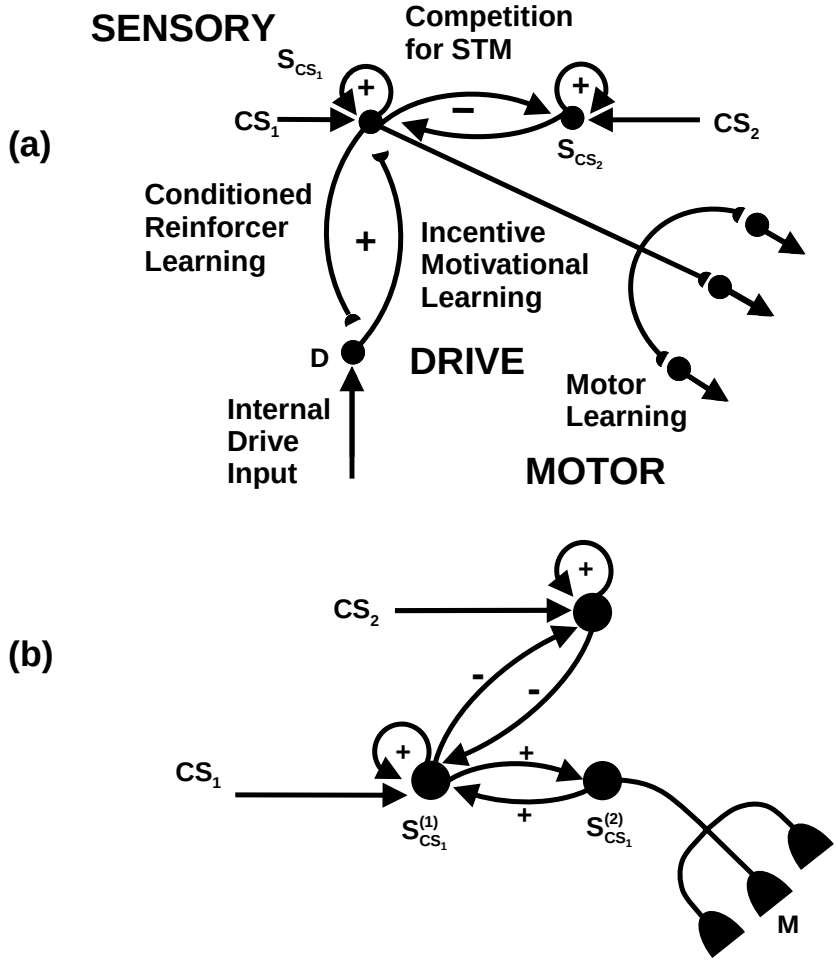


Figure 4. (a) The simplest CogEM model: Three types of interacting representations (sensory, drive, and motor) that control three types of learning (conditioned reinforcer, incentive motivational, and motor) help to explain many reinforcement learning data. Sensory representations *S* temporarily store internal representations of sensory events in working memory. Drive representations *D* are sites where reinforcing and homeostatic, or drive, cues converge to activate emotional responses. Motor representations *M* control the read-out of actions. Conditioned reinforcer learning enables sensory events to activate emotional reactions at drive representations. Incentive motivational learning enables emotions to generate a motivational set that biases the system to process information consistent with that emotion. Motor learning allows sensory and cognitive representations to generate actions. (b) In order to work well, a sensory representation *S* must have (at least) two successive stages, $S^{(1)}$ and $S^{(2)}$, so that sensory events cannot release actions that are motivationally inappropriate. See text for details.

Figure 4a summarizes the hypothesis that (at least) three types of internal representation interact during reinforcement learning: sensory and cognitive representations S, drive representations D, and motor representations M. Sensory representations S temporarily store internal representations of sensory events in working memory. Drive representations D are sites where reinforcing and homeostatic, or drive, cues converge to activate emotional responses. Motor representations M control the read-out of actions.

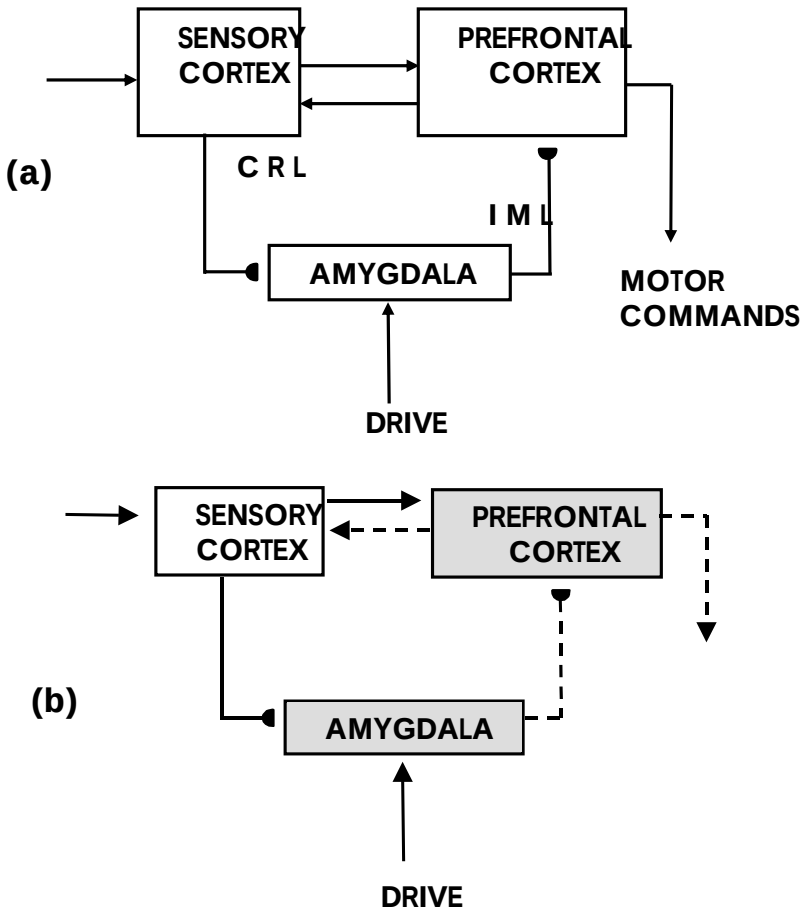


Figure 5. Anatomical interpretation of CogEM model. The two successive stages of a sensory representation *S* are interpreted to be in the appropriate sensory cortex (corresponds to $S^{(1)}$ in Figure 4b) and the prefrontal cortex, notably the orbitofrontal cortex (corresponds to $S^{(2)}$ in Figure 4b). The prefrontal stage requires motivational support from a drive representation *D* to be fully effective, in the form of feedback from the incentive motivational learning (IML) pathway. The amygdala is interpreted as one important part of a drive representation. Amydgala inputs to prefrontal cortex cause feedback to sensory cortex that selectively amplifies and focuses attention upon motivationally relevant sensory events. (b) When a drive representation like the amygdala gets depressed (gray box), diminished activation of its outputs in response to sensory events depresses motivational inputs to the prefrontal cortex in response to emotionally important events, and hereby attenuates motivationally-appropriate signals to and from the prefrontal cortex (dashed lines). As a result, motivationally irrelevant events are not attentionally suppressed, and prefrontally-mediated plans and actions are insufficently activated. See the text and Figure 6 for more details.

In particular, the S representations are thalamocortical representations of external events, including the object recognition categories that are learned by inferotemporal and prefrontal cortical interactions (Desimone, 1991; Gochin, Miller, Gross, and Gerstein, 1991; Harries and Perrett, 1991; Mishkin, Ungerleider, and Macko, 1983; Ungerleider and Mishkin, 1982), and that are modeled by ART. Sensory representations temporarily store internal representations of sensory events in short-term memory via recurrent on-center off-surround networks; see Figure 4. The D representations include hypothalamic and amygdala circuits at which reinforcing and homeostatic, or drive, cues converge to generate emotional reactions and motivational decisions (Aggleton, 1993; Bower, 1981; Davis, 1994; Gloor *et al*., 1982; Halgren *et al.*, 1978; LeDoux, 1993). The M representations include cortical and cerebellar circuits that control discrete adaptive responses (Evarts, 1973; Ito, 1984; Kalaska *et al.*, 1989; Thompson, 1988). More complete models of the internal structure of these several types of representations are developed elsewhere (e.g., Brown, Bullock, and Grossberg, 2004; Bullock, Cisek, and Grossberg, 1998; Carpenter and Grossberg, 1991; Contreras-Vidal, Grossberg, and Bullock, 1997; Fiala, Grossberg, and Bullock, 1996; Grossberg, 1987; Grossberg and Merrill, 1996; Grossberg and Schmajuk, 1987; Raizada and Grossberg, 2003). The model even in this simple form has successfully learned to control motivated behaviors in mobile robots (e.g., Baloch and Waxman, 1991; Chang and Gaudiano, 1998; Gaudiano and Chang, 1997; Gaudiano, Zalama, Chang, and Lopez-Coronado, 1996).

Three types of learning take place among these representations: Conditioned reinforcer learning (CRL) enables sensory events to activate emotional reactions at drive representations. Incentive motivational learning (IML) enables emotions to generate a motivational set that biases the system to process cognitive information consistent with that emotion. Motor learning allows sensory and cognitive representations to generate actions.

In particular, learning within the S $\rightarrow$ D conditioned reinforcer pathways converts a CS into a reinforcer when it activates its sensory representation S just before the drive representation D is activated by an unconditioned stimulus (US), or other previously conditioned reinforcer CSs. The ability of the CS to subsequently activate D via this learned pathway is one of its key properties as a conditioned reinforcer. As these S $\rightarrow$ D associations are being formed, incentive motivational learning within the D $\rightarrow$ S incentive motivational pathways also occurs, due to the same pairing of CS and US. Incentive motivational learning enables an activated drive representation D to prime, or modulate, the sensory representations S of all cues, including the CSs, that have consistently been correlated with it. That is how activating D generates a "motivational set": it primes all of the sensory and cognitive representations that have been associated with that drive in the past. These incentive motivational signals are a type of motivationally-biased attention. The S $\rightarrow$ M motor, or habit, learning enables the sensorimotor maps, vectors, and gains that are involved in sensory-motor control to be adaptively calibrated, thereby enabling a CS to read-out correctly calibrated movements.

Taken together, these processes control the learning and recognition of sensory and cognitive memories, which are often classified as part of a "declarative memory" system (Mishkin, 1982, 1993; Squire and Cohen, 1984); and the performance of learned motor skills, which are often classified as part of a "procedural memory" system (Gilbert and Thatch, 1977; Ito, 1984; Thompson, 1988).

6.2. Motivationally Appropriate Actions using Multiple Stages of Sensory Representation

In order to generate only motivationally appropriate behaviors, the circuit in Figure 4a needs to

have two successive stages of sensory processing which are interpreted to occur in a sensory cortex and a prefrontal cortex to which it projects (Figure 5a). If only a single sensory cortex were used (that is, lump the sensory and prefrontal cortex together, as in Figure 4a), then, after a reinforcing cue activates a sensory representation S, it can activate a motor representation M at the same time that it also sends conditioned reinforcer signals to a drive representation D such as the amygdala. Then a motor response could be initiated before the sensory representation receives incentive motivational feedback to determine whether the sensory cue *should* generate a response at that time. For example, eating behavior could be initiated before the network could determine if it was hungry. This deficiency is corrected by using a sensory cortex and its prefrontal cortical projection, as in Figure 4b and 5a. Here, the various sensory cortices play the role of the first cortical stage of the sensory representations, the (orbital) prefrontal cortex plays the role of the second cortical stage of the sensory representations, and the amygdala and related structures play the role of the drive representations.

This two-stage sensory representation overcomes the problem just mentioned by embodying a key property of drive representations. Each drive representation D obeys a *polyvalent* constraint whereby it can generate large incentive motivational output signals to sensory representations S only if it gets a sufficiently large primary or conditioned reinforcer input at the same time that it gets a sufficiently large internal drive input. The internal drive input designates whether an internal drive, such as hunger, thirst, sex, etc. is high and in need of satisfaction. Different drive representations exist to represent these distinct internal homeostatic states. The polyvalent constraint means that a drive representation cannot fire vigorously unless it simultaneously receives a sufficiently large external sensory input and internal drive input. Due to the polyvalent constraint at the drive representation, an external reinforcing cue cannot activate strong incentive motivation, and with it action, to satisfy a drive that is already satisfied, because the drive input would be too small.

By imposing a similar polyvalent constraint on prefrontal cortical cells, the sensory representations cannot trigger an action unless they simultaneously receive sensory input from the corresponding sensory cortex and incentive motivational input from a drive representation. Thus, the polyvalent constraint on prefrontal firing prevents this region it from triggering an action until it gets incentive feedback from a motivationally-consistent drive representation (Grossberg, 1971, 1987).

More specifically, presentation of a given cue, or CS, activates the first stage of its sensory representation (in sensory cortex); see $S_{CS}^{(1)}$ in Figure 4b. This activation is stored in short-term memory using positive feedback pathways from the sensory representation to itself. The stored activity generates output signals to all the drive representations with which the sensory representation is linked, as well as to the second stage $S_{CS}^{(2)}$ of the sensory representation (in prefrontal cortex; see Figure 5a). The second stage of the sensory representation obeys a polyvalent constraint: It cannot fire while the CS is stored in short-term memory unless it receives converging signals from the first sensory stage and from a drive representation.

This polyvalent constraint solves the problem of motivationally-inappropriate release of actions as follows: Suppose that the first stage of the sensory representation sends a large primary or secondary reinforcing signal to a drive representation at a time when the drive representation is also receiving a sufficiently large drive input. Then the polyvalent constraint of the drive representation is satisfied and the drive representation can fire. In other words, the drive representation can fire when the drive is not yet satisfied and sensory cues are available that predict drive satisfaction. All the drive representations that are active at that time compete among

themselves to allow the most active one—the one that represents the best combination of sensory and drive information at that moment—to fire. If the winning drive representation has a strong learned incentive motivational pathway to the second stage of the active sensory representation, then the polyvalent constraint of the second stage is satisfied, and it can generate output signals. In summary, by making the final stages of both the sensory and the drive representations polyvalent, the S → M motor pathways are activated only if the S → D → S feedback pathway can get sufficiently activated. In other words, the network generates a strong conditioned response only if it receives enough motivational support.

6.3. Motivated Attention and Blocking

Positive feedback from the prefrontal cortex to its sensory cortex also exists, as proposed by ART in order to stabilize rapid neocortical learning; see Figures 4b and 5. Such positive feedback can direct attention to motivationally salient sensory events, and to select motivationally appropriate responses. (Reynolds and Chelazzi (2004) provide an excellent review of attentional modulation of visual processing.) In other words, attentional feedback can help to explain how attentional blocking can occur, whereby the sensory representations of unattended and irrelevant sensory cues can be suppressed and thereby prevented from being learned (Grossberg and Levine, 1987; Pavlov, 1927). How does such cortical feedback enable attentional blocking to occur?

As noted above, model prefrontal cortical cells can fire only if a drive representation with which it is associated receives strong incentive motivational inputs from a drive representation. As a result, positive feedback from the prefrontal cortex to the sensory cortex amplifies only those active sensory representations that are motivationally prepotent in the present context. This amplification of activity enables these sensory representations to attentionally block less salient representations via the recurrent lateral inhibitory connections that exist among the sensory representations. These inhibitory connections are part of the recurrent on-center off-surround networks that store sensory cues in short-term memory without a loss of contrast sensitivity (Grossberg, 1973, 1980, 1982a).

6.4. Interactions between Sensory Cortices, Amygdala, and Orbital Prefrontal Cortex

The circuit in Figure 4a may, in principle, be replicated at multiple stages of thalamocortical and corticocortical processing of sensory events. For example, the first sensory stage may be a thalamic stage, and the second sensory stage may be a neocortical stage, as in the data of LeDoux (1993). The circuit in Figure 5a is consistent with Figure 6, which is adapted from Barbas (1995), who noted that many different types of sensory cortex, including visual, somatosensory, auditory, gustatory, and olfactory cortex, are connected to both the amygdala (and other emotional centers) and to the prefrontal cortex, and that the amygdala also sends a strong projection to the prefrontal cortex. We interpret this anatomy in terms of the model circuit in Figure 4b where, as noted above, the various sensory cortices are the first cortical stage of the sensory representations, the (orbital) prefrontal cortex is the second cortical stage of the sensory representations, and the amygdala and related structures are the drive representations.

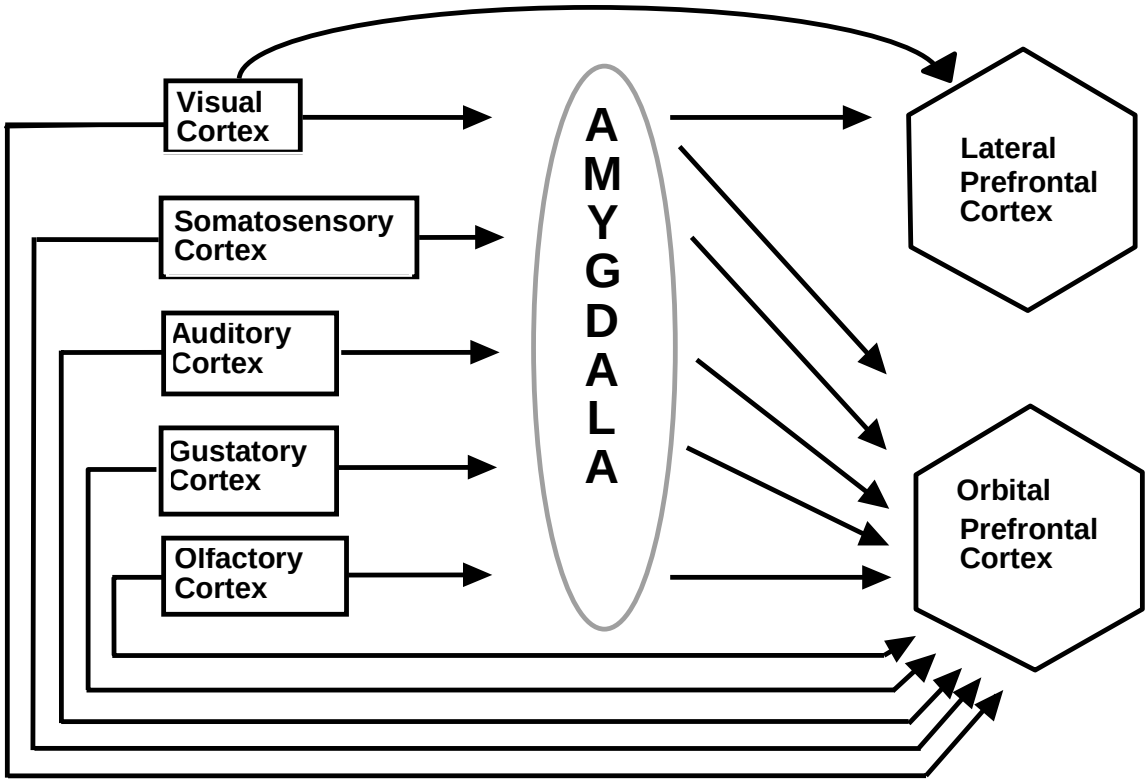


Figure 6. Orbital prefrontal cortex receives projections from sensory cortex (visual, somatosensory, auditory, gustatory, and olfactory) and from the amygdala, which also receives inputs from the same sensory cortices. [Reprinted with permission from Barbas (1995).]

The following properties of Figure 5a are consistent with this anatomical interpretation: The amygdala, and related structures, has been identified in both animals and humans to be a brain region that is involved in learning and eliciting memories of experiences with strong emotional significance (Aggleton, 1993; Davis, 1994; Gloor *et al.*, 1982; Halgren *et al.*, 1978; LeDoux, 1993). The orbitofrontal cortex is known to be a major projection area of the ventral, or object-processing cortical visual stream (Barbas, 1995; Fulton, 1950; Fuster, 1989; Rolls, 1998; Wilson *et al*., 1993). Cells in the orbitofrontal cortex are sensitive to the reward associations of sensory cues, as well as to how satiated the corresponding drive is at any time (e.g., Mishkin and Aggleton, 1981; Rolls, 1998). The feedback between the prefrontal and sensory cortical stages may be interpreted as an example of the ubiquitous positive feedback that occurs between cortical regions (Felleman and Van Essen, 1991; Macchi and Rinvik, 1976; Sillito *et al*., 1994; Tsumoto, Creutzfeldt, and Legéndy, 1978; van Essen and Maunsell, 1983), including prefrontal and sensory cortices. Finally, the model is also consistent with data suggesting that the ventral prefrontal cortex and the amygdala are involved in the process by which responses are selected on the basis of their emotional valence and success in achieving rewards (Damasio *et al.*, 1991; Passingham, 1997). In further support of the CogEM model hypothesis that the prefrontal sensory representation gates the release of properly motivated actions, Fuster (1989) has concluded from studies of monkeys that the orbital prefrontal cortex helps to suppress inappropriate responses. These monkey data are consistent with clinical evidence that patients

with injury to orbital prefrontal cortex tend to behave in an inappropriate manner (Blumer and Benson, 1975; Liddle, 1994).

As noted in Figure 5a, amydgala inputs to prefrontal cortex cause feedback to sensory cortex that selectively amplifies and focuses attention upon motivationally relevant sensory events. When a drive representation like the amygdala cannot fire at normal levels, as in Figure 5b (gray box), its diminished outputs in response to sensory events depresses motivational inputs to the prefrontal cortex in response to emotionally important events. The polyvalent constraint at the prefrontal cortex is therefore no longer satisfied, and the prefrontal cortex cannot fire adequately. Motivationally-appropriate signals to and from the prefrontal cortex (dashed lines) are hereby attenuated. As a result, motivationally irrelevant events are not attentionally suppressed, and prefrontally-mediated plans and actions are insufficently activated.

The CogEM model proposes that emotional centers like the amygdala can experience such a reduced overall output, which can thereby cause the types of problems with motivated attention and action that has just been summarized. This condition is referred to as a form of "emotional depression." Although emotionally-charged outputs from amygdala may be reduced, such reduction is not herein identified with the full-blown clinical disease of depression.

How does such emotional depression arise in the model? How this occurs can be seen by noting that emotional centers are often organized into *opponent* affective processes, such as fear and relief (Grossberg, 1984b, 2000b). These opponent-processing emotional circuits are called *gated dipoles* for reasons described below. The response amplitude and sensitivity to external and internal inputs of these opponent-processing emotional circuits are calibrated by an arousal level and chemical transmitters that slowly inactivate, or habituate, in an activity-dependent way. These opponent processes exhibit an Inverted-U whereby their outputs may become depressed if the arousal level is chosen too large or too small (Figure 7). Underaroused and overaroused depression can be distinguished clinically by their parametric properties, some of which are summarized in Figure 7 and proved mathematically in the Appendix. The iSTART model proposes that some symptoms of autism are due to underaroused depression and the way in which this condition interacts with other circuits, notably cognitive and motor circuits, throughout the brain. In particular, if the amygdala experiences underaroused depression, then this deficiency could ramify throughout the brain in the manner schematically shown in Figure 5b.

6.5. Opponent Processing: Inverted-U, Antagonistic Rebound, and Attentional Perseveration

Given this background, let us now consider in more detail how opponent emotions, like fear and relief, may be organized in the brain, and how they may become depressed. ART predicts that such opponent emotions are a special case of a more general brain design for opponent processing, including opponent perceptual features like red and green colors, or downward and upward motions, or horizontal and vertical orientations. All of these different examples have the property of generating *antagonistic rebounds* whereby, say, offset of a sustained fearful cue can elicit a wave of relief, or removal of a desired food can elicit a wave of frustration, or offset of a sustained red image can yield a green aftereffect, or offset of a sustained downward motion of water can yield an upward motion aftereffect, or offset of a sustained image with radial spokes of a wheel can yield an aftereffect of concentric circles, and so on. In all of these cases, there are ON and OFF channels that can experience an antagonistic rebound. These opponent rebounds are predicted to play a key role in controlling ART reset and search, as discussed above, as well as in rebalancing sensory, cognitive, emotional, and motoric representations in response to rapidly

changing environmental inputs. They have been used to explain a wide variety of behaviors about animal and human cognitive-emotional learning and decision-making (e.g., Grossberg, 1972b, 1980, 1984a, 1984b, 2000b, Grossberg and Gutowski, 1987; Grossberg and Schmajuk, 1987).

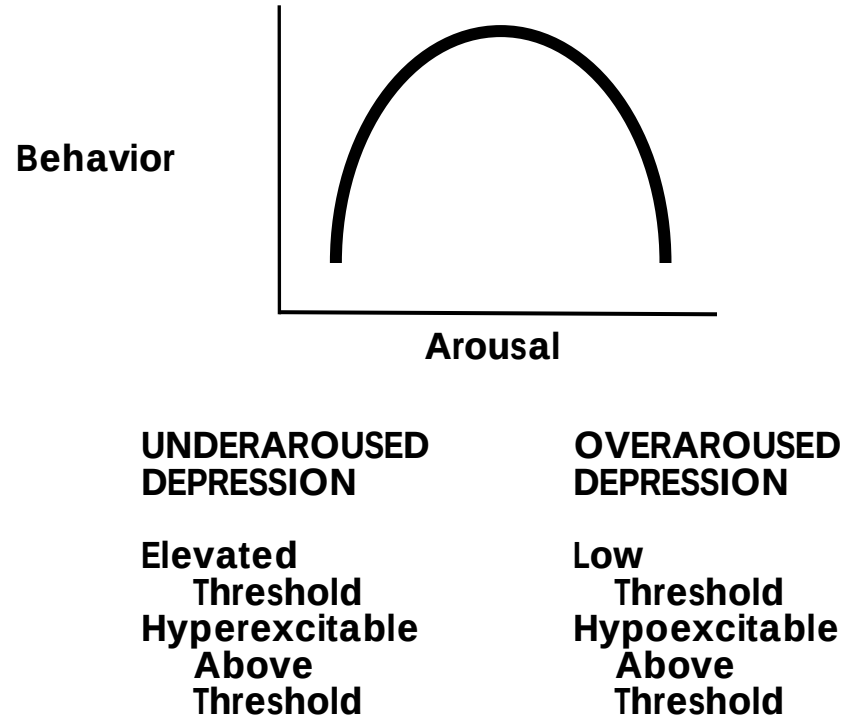


Figure 7. Gated dipole opponent processes exhibit an Inverted-U behavioral response as a function of arousal level, with underaroused and overaroused depressive syndromes occurring at the two ends of the Inverted-U. See text for details. [Reprinted with permission from Grossberg (2000b).]

Such opponent processing circuits exhibit a Golden Mean of optimal behavior at an intermediate arousal level, as noted in Figure 7. For larger or smaller levels of arousal, behavior deteriorates in different ways, thereby giving rise to an Inverted-*U* in network performance as a function of its arousal level. In different parts of the brain, the arousal source may differ. In general, a baseline of arousal means an input that is *tonically* on (or internally generated and active during an interval of fast phasic inputs) and which is received equally by both the ON and OFF channels of the dipole.

Both the Inverted-*U* and the antagonistic rebound are the result of *habituative* chemical transmitters, or transmitters that are released in an activity-dependent way faster than they can recover. These transmitters exist in the opponent channels, where they multiply, or gate, the signals on their way to the opponent, or competitive, processing stage. Due to the factors that are summarized in Figure 8a and proved in the Appendix, when arousal is too small, such an opponent process experiences an elevated response threshold in response to an ON channel input, since there is not enough tonic arousal to enable outputs to occur in response to normal phasic input levels. Paradoxically, such an underaroused circuit also gives rise to *hyper*excitable,

or larger than normal, responses to increments in the ON input that exceed this elevated threshold. This is true because the arousal input, when it is gated by the habituative transmitter, acts like a gain that divides cell activation. Division by a smaller arousal level causes larger suprathreshold cell activations. When arousal is too large, the opponent process experiences a low behavioral threshold, since there is plenty of tonic arousal to boost the effect of phasic inputs. Paradoxically, an overaroused dipole gives rise to *hypo*excitable, or smaller than normal, outputs, in response to increments in the ON input that exceed the reduced threshold, because an abnormally large arousal divides the cell activation. Due to these properties, an *in*crease in arousal can *de*crease the sensitivity of an underaroused opponent process of this kind, and can bring it into the normal behavioral range.

This opponent processing model is called a "gated dipole" because habituative transmitters "gate," or multiply, signal processing in each of the channels of the opponent "dipole." Due to the Inverted-*U* property of a gated dipole, which is mathematically proved in the Appendix, a suitable pharmacological "up" should reduce the supra-threshold hypersensitivity of patients with underaroused dipole circuits, as is assumed herein for individuals with autism.

Figure 8a defines the simplest gated dipole using a feedforward circuit. In general, gated dipoles are defined by feedback circuits in order to store their operating levels in short-term memory and thereby, for example, to maintain a steady motivational baseline while an entire action is being carried out. Figure 8b illustrates a feedback gated dipole that is called a READ (REcurrent Associative Dipole) circuit. In includes associative pathways whereby a sensory representation S_k can learn to activate either the ON or OFF channels of the dipole via adaptive weights w_k. Retrograde dendritic spikes (the region below x_7 and x_8 in Figure 8b) dissociate the read-out of previous emotional memories from the read-in of new emotional memories, which enables learning to remain sensitive to changing reward contingencies. See Grossberg and Schmajuk (1987) for computer simulations of how a READ circuit can learn and remember stable emotional memories for many years until they are actively erased by disconfirmed sensory or cognitive expectations.

Sensory and cognitive representations, as well as emotional representations, can be organized into opponent channels with habituative ON and OFF cells. ART illustrates how an unexpected event can trigger a burst, or sudden increment, of nonspecific arousal (see Figure 2c). When such an arousal burst is received on top of the baseline tonic arousal input of a normal dipole, it can cause an antagonistic rebound of activity in the OFF channel. In other words, the sensory, cognitive, or emotional hypothesis that is represented in a dipole's activity can be disconfirmed by an unexpected event. In such a normal dipole, an unexpected event can hereby reset ongoing processing and lead to a shift of attention. In an underaroused dipole, by contrast, an unexpected event can cause a paradoxical amplification of activity in the ON channel of the dipole, instead of the more normal antagonistic rebound of activity in the OFF channel. Thus, instead of resetting the dipole in response to an unexpected event, an unexpected event can instead cause the dipole to maintain, indeed enhance, the activity in its currently active channels. Such an enhancement can result in perseveration of attention on a given object of event, rather than flexible disengagement and shifting of attention. Split attention, shifting attention, and joint attention, which requires a flexible shift of the balance of split attention from an object of social value, such as a mother's gaze, to another object and back again, may thus all be impaired in such an underaroused model. If, however, an arousal burst is sufficiently strong, then an unusually intense antagonistic rebound can be caused. See the Appendix for a mathematical

proof of these properties. These underaroused properties are observed among individuals with autism.

A gated dipole's perseverative and rebound properties emerge through interactions across the entire gated dipole circuit. They cannot be understood just by looking at the pharmacology or neurophysiology of individual cells within the circuit. When their effects ramify throughout the sensory and prefrontal cortices with which they interact, as in Figures 4 and 5, they can lead to a number of clinical symptoms.

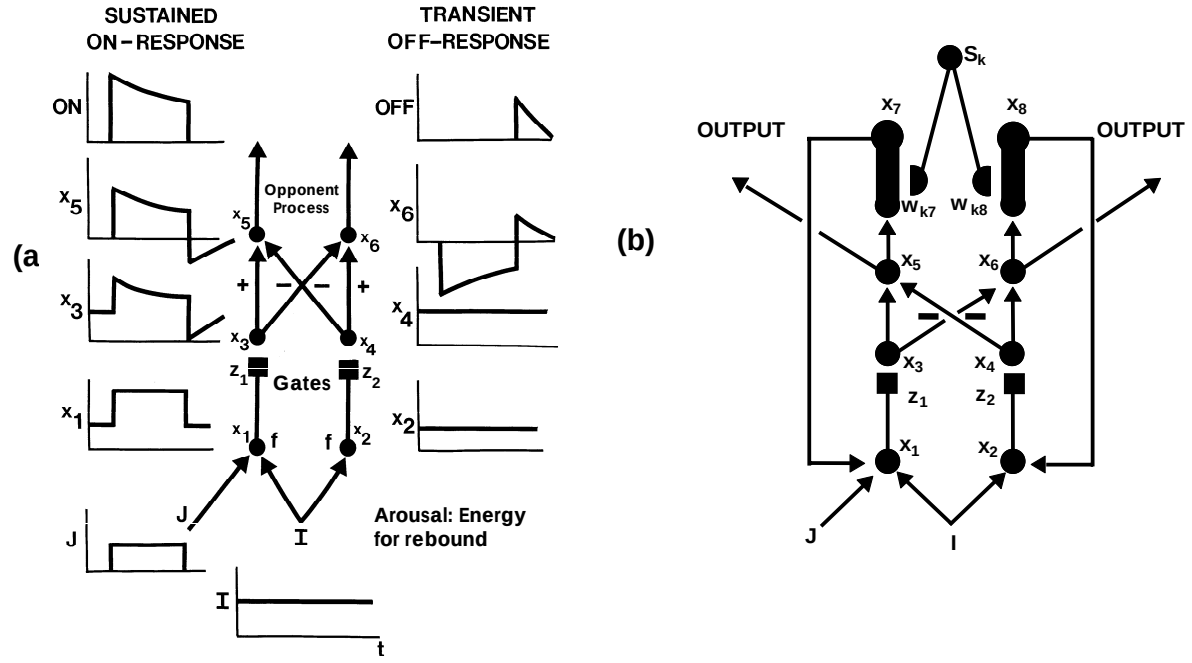


Figure 8. (a) A gated dipole opponent process can generate habituative ON responses and transient OFF rebounds in response to the phasic onset and offset, respectively, of the input J to its ON channel. Term I delivers tonically active nonspecific arousal that energizes antagonistic rebounds when the phasic input J shuts off. Terms z_1 and z_2 are the habituative transmitter gates, or depressing synapses, in the ON and OFF channels; see the square synaptic symbols. They multiply, or gate, signals $f(x_1)$ and $f(x_2)$ that are derived from the ON and OFF activities x_1 and x_2, respectively, before the gated signals $f(x_1)z_1$ and $f(x_2)z_2$ excite the activities x_3 and x_4, respectively. The habituative gates convert the step-on-baseline activity pattern x_1 into the overshoot-habituation-undershoot-habituation pattern at activity x_3. Next, the opponent interaction works; namely, the baseline activity x_4 in the OFF channel due to the arousal I is subtracted from the habituative ON activity x_3 to yield x_5. When activity x_5 is thresholded to generate an ON output signal, it has an initial overshoot of activation, followed by habituation. When the signs of excitation and inhibition are reversed in the OFF channel, the activity x_6 is caused. The antagonistic rebound in the OFF output is generated by thresholding x_6. The antagonistic rebound is thus derived from the mirror-image of excitation and inhibition of the undershoot-habituation part of the ON channel activity at x_5. [Reprinted with permission from Grossberg (2000b).]; (b) A READ (Recurrent Associative Dipole) circuit. The READ circuit is a gated dipole with excitatory feedback pathways between activities x_7 and x_1, and activities x_8 and x_2. Feedback enables the READ circuit to maintain a stable motivational baseline to support an ongoing motivated behavior. A sensory representation S_k sends conditionable signals to the READ circuit that are gated by conditioned reinforcer adaptive weights, or long term memory traces, w_{k7} and w_{k8} to the ON and OFF channels, respectively. Read-out of previously learned adaptive weights is dissociated from read-in of new values of the learned weights. New weight learning is generated by teaching signals from the ON or OFF channel that wins the opponent competition. The combination of recurrent feedback and associative dissociation enables the adaptive weights to avoid learning baseline noise, maintain sensitivity to the relative balance of ON and OFF channel conditioning through time, and preserve their learned memories until they are disconfirmed by new learning contingencies. [Reprinted with permission from Grossberg and Schmajuk (1987).]

6.6. When Hypervigilant Cognitive Learning Modulates Underaroused Emotional Circuits

Let us suppose that certain individuals with autism have a variety of underaroused emotional as well as sensory and cognitive opponent representations (not necessarily all sensory and cognitive representations, however). Consider what can happen when this property is combined with hyperspecific and hypervigilant cognitive learning. That is, consider how an underaroused CogEM model interacts with a hypervigilant ART model. In this combined system, various formal symptoms emerge that strikingly resemble behavioral properties of individuals with autism.

For example, suppose that positive affect motivates a learned action. As noted in Section 6.5, underaroused emotional and sensory dipoles can exhibit a paradoxical enhancement of their ON channels when a nonspecific arousal burst is triggered by an unexpected event. How can such paradoxical enhancements be caused in an individual with autism? Imagine that such an individual inspects an object closely in slightly different ways. The hyperspecific top-down expectations of such an individual may cause mismatches with the different views of the object. These mismatches can cause nonspecific arousal bursts, which can cause enhanced ON channel responses. In particular, enhanced positive affect in underaroused emotional dipoles can lead to enhanced storage of the object representation in sensory and cognitive dipoles. A persistent and self-reinforcing perseverative behavior can result, which might manifest itself in persistently inspecting the same object, over-and-over, from slightly different perspectives.

Suppose, however, that an arousal burst is caused which is significantly larger, say due to a larger mismatch of a presently active hyperspecific category prototype with a very different and unexpected event in the world. Then an unusually intense, and negative, antagonistic rebound can be caused. Thus, novel experiences can be highly aversive when hyperspecific categories mismatch them and suddenly generate a burst of arousal to underaroused emotional dipoles. These negative rebounds may be one reason why individuals with autism are prone to experiencing severe negative reactions to unanticipated events.

When one considers the plight of any individual who combines these two mechanisms, it becomes clear that one coping strategy is to avoid the type of novelty that will cause unbearable negative rebounds. The other side of the coin is perseveration on small details of the environment. Such a combination of properties may help to understand the autistic "need for sameness" in many situations, and the fact that behavioral decompensation often occurs in response to what a normal person would view as relatively minor variations in routines.

6.7. Impoverishment of Motivated Goals, Intentions, and Theory of Mind

The most immediate effect of a depressed response in the outputs of emotion-representing areas is flat affect—that is, a reduced output from emotional centers—although how this is understood must be carefully evaluated, as indicated in the preceding discussion of underaroused and overaroused depression. Flat affect may cause an inability to represent others' beliefs and intentions, in the sense that all mental states that depend upon interpreting one's own emotional state, or the emotional states of others, can be diminished. Such a deficiency can cause major difficulties in social communication. It happens in the CogEM model of Figures 4 and 5 because emotionally charged sensory inputs, such as the expressions on other people's faces or their tone of voice, will activate the appropriate part of a temporal cortex but may not elicit an appropriate emotional output from the amygdala and related emotion-representing circuits via conditioned reinforcer pathways.

A problem with the setting of motivationally directed goals and intentions can then indirectly arise. This happens in the model because the depressed response of the emotional

representations, in brain areas like the amygdala, depresses the incentive motivational signals that would normally activate the prefrontal cortex in response to motivationally salient events (Figure 5b). As a result, the prefrontal cortex will not be adequately activated, and a hypofrontal condition can emerge. Due to this hypofrontality, the working memory representations and plans that are ordinarily formed within the prefrontal cortex will be degraded, so social goals and plans will not form in a normal fashion. The combination of depressed affective responses to environmental and internally generated cues, combined with insufficient motivational support for emotionally-appropriate plans and actions, helps to explain why individuals with autism are said to be without a Theory of Mind; see Section 9.

Given such a hypofrontal response, top-down signals from the prefrontal cortex to the sensory cortices will also be reduced or eliminated (Figure 5b). As a result, the sensory representations will not be able to use these top-down signals to organize information processing according to its emotional meaning or motivational goals.

7. Adaptive Timing Model

7.1. Adaptively Timed Learning, Motivation, Attention, and Action

The above discussion illustrates one aspect of a major conceptual dichotomy that is often used in research about normal and amnesic learning and memory. This dichotomy concerns the distinction between processes that are variously called declarative memory and procedural memory, knowing that and knowing how, memory and habit, or memory with record and memory without record (Bruner, 1969; Mishkin, 1982, 1993; Ryle, 1949; Squire and Cohen, 1984). The amnesic patient HM exemplified this distinction by learning and remembering motor skills like assembly of the Tower of Hanoi without being able to recall having done so (Bruner, 1969; Cohen and Squire, 1980; Mishkin, 1982; Ryle, 1949; Scoville and Milner, 1957; Squire and Cohen, 1984). HM's surgical lesion included extensive parts of the hippocampal formation and amygdala. Subsequent animal studies have shown that damage to the hippocampal formation (Ammon's horn, dentate gyrus, subiculum, fornix) and the parahippocampal region (entorhinal, perirhinal, and parahippocampal cortices) can reproduce analogous amnesic symptoms (Mishkin, 1978; Squire and Zola-Morgan, 1991). These results implicate this aggregate hippocampal system in the processes that regulate declarative memory, or "knowing that". Such processes support a competence for learning recognition categories and being able to flexibly access them in a task-specific way (Eichenbaum, Otto, and Cohen, 1994). The discussion of ART above is about declarative memory, particularly about the learning of recognition categories, and involves predicted interactions between cortical and hippocampal representations. Indeed, as noted in Section 5.2, ART exhibits a constellation of formal symptoms that strikingly resemble symptoms of medial temporal amnesia when its orienting system is lesioned (Carpenter and Grossberg, 1993).

A parallel line of research has implicated the cerebellum in the processing of procedural memory, or "knowing how". The cerebellum is an essential circuit for conditioning discrete adaptive responses during eye movements, arm movements, nictitating membrane movements, and jaw movements (Ebner and Bloedel, 1981; Gilbert and Thach, 1977; Ito, 1984; Lisberger, 1988; Optican and Robinson, 1980; Thompson, 1988; Thompson *et al*., 1984, 1987). Models of cerebellar learning have been developed over the years to help explain these motor conditioning data (Albus, 1971; Bullock, Fiala, and Grossberg, 1994; Fiala, Grossberg, and Bullock, 1996; Fujita, 1982a, 1982b; Grossberg, 1969b, 1972c; Grossberg and Kuperstein, 1986; Ito, 1984; Lisberger, 1988; Marr, 1969).

A key property of cerebellar learning is that it is *adaptively timed*, so that learned

responses are emitted at times that are appropriate within the environmental constraints of the learning paradigm. Cognitive-emotional learning is also adaptively timed, so that motivated attention can be maintained on salient goal objects for the necessary amount of time to carry out goal-directed actions.

Many goal objects may be delayed subsequent to the actions that elicit them, or the environmental events that signal their subsequent arrival. Humans and many animal species can learn to wait for the anticipated arrival of a delayed goal object, even though its time of occurrence can vary from situation to situation. Such behavioral timing is important in the lives of animals which can explore their environments for novel sources of gratification. On the one hand, if an animal or human could not inhibit its exploratory behavior, then it could starve to death by restlessly moving from place to place, unable to remain in one place long enough to obtain delayed rewards there, such as food. On the other hand, if an animal inhibited its exploratory behavior for too long while waiting for an expected source of rewards, such as food, to materialize, then it could starve to death if food is not, after all, forthcoming.

Thus, the survival of a human or animal may depend on its ability to accurately time the delay of a goal object based upon its previous experiences in a given situation. Such human or animal needs to balance between its exploratory behavior, which may discover novel sources of reward, and its consummatory behavior, which may acquire expected sources of reward. To effectively control this balance, the human or animal needs to be able to suppress its exploratory behavior and focus its attention upon an expected source of reward at around the time that the expected delay transpires for acquiring this reward.

7.2. Adaptively Timed Gating

To illustrate the sort of timing that is intended here, suppose that an animal typically receives food from a food magazine 2 seconds after pushing a lever, and that the animal orients to the food magazine right after pushing the lever. When the animal inspects the food magazine, it perceives the non-occurrence of food during the subsequent 2 seconds. These non-occurrences disconfirm the animal's sensory expectation that food will appear in the magazine. Because the perceptual processing cycle that processes this sensory information occurs at a much faster rate than 2 seconds, it can compute this sensory disconfirmation many times before the 2 second delay has elapsed.

The core issue is: What spares the animal from erroneously reacting to these *expected non-occurrences* of food during the first 2 second as predictive failures? Why does the animal not immediately become so frustrated by the non-occurrence of food that it shifts its attentional focus and releases exploratory behavior aimed at finding food somewhere else? Alternatively, if the animal does wait, but food does not appear after the 2 seconds have elapsed, then why does the animal then react to the *unexpected non-occurrence* of food by becoming frustrated, resetting it working memory, shifting its attention, and releasing exploratory behavior?

Any solution to this problem needs to account for the fact that the processing of registering sensory mismatches or matches is not itself inhibited: If the food happened to appear earlier than expected, the animal could still perceive it and eat. Instead, the *effects* of these sensory mismatches upon reinforcement, attention, and exploration are somehow inhibited, or *gated* off. That is, a primary role of such an adaptive timing mechanism seems to be to inhibit, or gate, the process whereby a disconfirmed expectation would otherwise negatively reinforce previous consummatory behavior, reset working memory, shift attention, and release exploratory behavior.

In summary, unless motivated attention and action are both adaptively timed, an animal

or human could be condemned to either emit premature goal-oriented responses (as occurs in individuals with autism; see Section 3), or to generate maladaptive orienting and exploratory movements in any situation wherein a goal object does not immediately appear. Adaptively timed learning enables both attention and action to be appropriately timed to generate adaptive behavior in each environment.

Evidence for adaptive timing has been found during many different types of reinforcement learning. For example, during classical conditioning, a conditioned stimulus (CS) such as a tone or light, when paired with an unconditioned stimulus (US) such as a shock, can learn to generate conditioned responses (CR), such as fear or limb withdrawal, that were originally elicited only by the US. Such learning is optimal at a range of positive interstimulus intervals (ISI) between the CS and US that are characteristic of the animal and the task, and is greatly attenuated at zero ISI and long ISIs. Within this range, learned responses are timed to match the statistics of the learning environment (Smith, 1968). Although the amygdala has been identified as a primary site in the expression of emotion and stimulus-reward association (Aggleton, 1993), as summarized in Figures 5 and 6, the hippocampal formation has been implicated in the adaptively timed processing of cognitive-emotional interactions. For example, Thompson *et al*. (1987) distinguished two types of learning that go on during conditioning of the rabbit Nictitating Membrane Response: Adaptively timed "conditioned fear" learning that is linked to the hippocampus, and adaptively timed "learning of the discrete adaptive response" that is linked to the cerebellum.

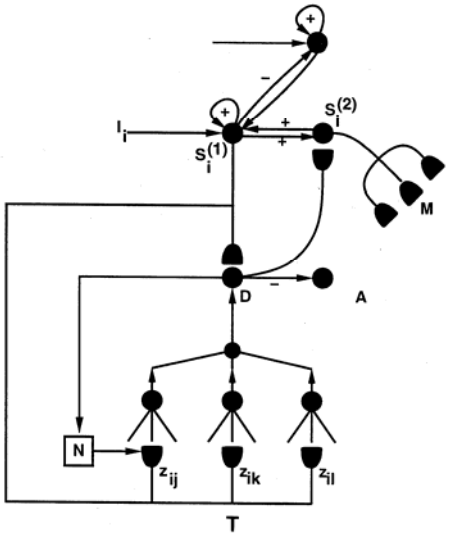


Figure 9. The simplest version of the START model. Adaptively timed learning maintains motivated attention at the same time that it inhibits activation of the orienting system. See text for details. [Reprinted with permission from Grossberg and Merrill (1992).]

A synthesis of the ART and CogEM models, called the START model, for Spectrally Timed ART model (Fiala, Grossberg, and Bullock, 1996; Grossberg and Merrill, 1992, 1996; Grossberg and Schmajuk, 1987), proposes a unified explanation of why both the hippocampal system and the cerebellum may need adaptive timing circuits for their normal functioning (Figures 9 and 10). Appendix B summarizes the equations of the START model that was simulated by Grossberg and Merrill (1992). The START model predicts how motivational mechanisms within

w_{ij} w_{ik} w_{il}

the amygdala, and related emotion-representing brain areas, can rapidly draw motivated attention to salient cues. This can happen via a cognitive-emotional resonance within CogEM feedback circuits between sensory representations *S* and drive representations *D* (see Figure 9). Once these salient cue representations are selected and activated, what prevents the actions that they control from being prematurely released via the CogEM circuit in Figures 4b and 5?

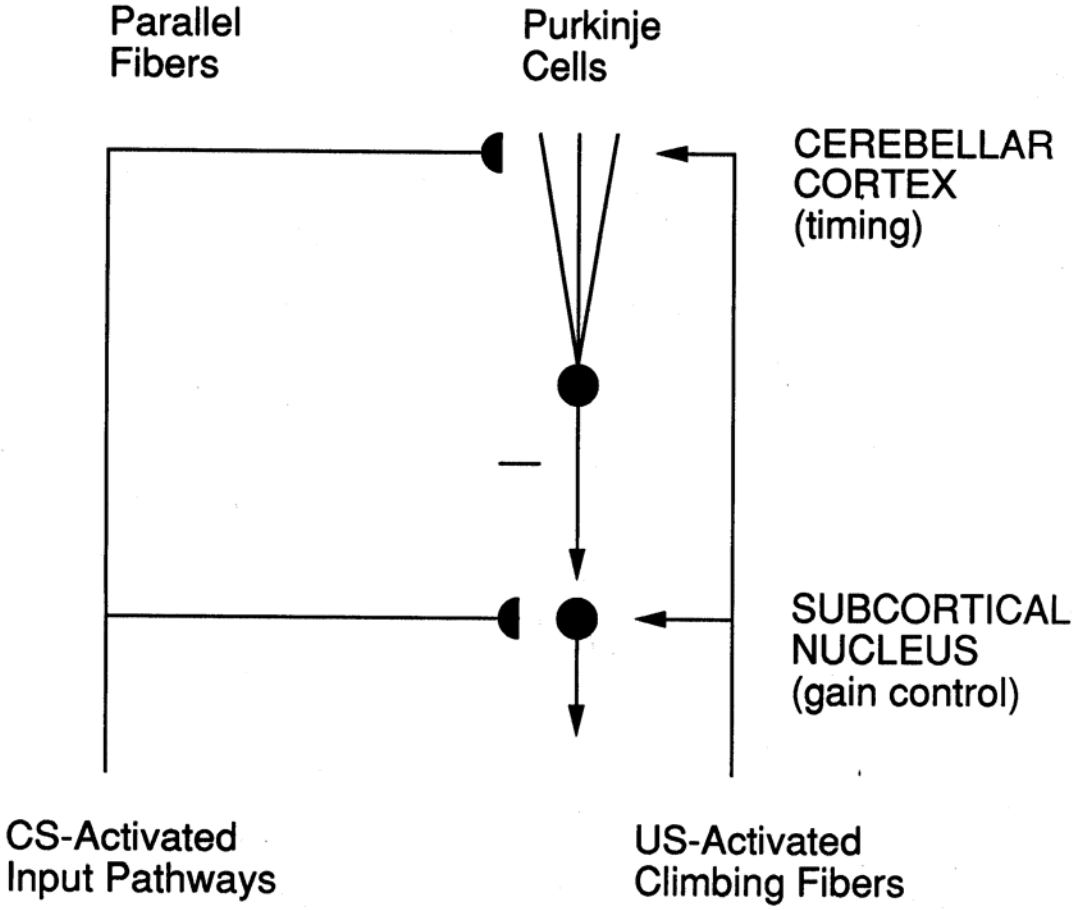


Figure 10. Circuit for adaptively timed cerebellar learning. Adaptively timed Long Term Depression at Purkinje cells depresses the level of tonic inhibitory firing of these cells to cerebellar nuclei, thereby disinhibiting nuclear cells and allowing them to expressed their learned gains in an adaptively timed way. See text for details. [Reprinted with permission from Grossberg and Merrill (1996).]

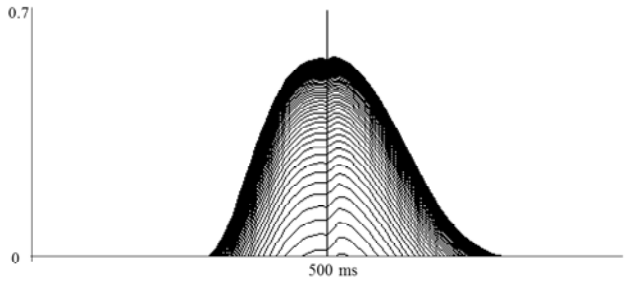
7.3. Spectral Timing in Cerebellum and Hippocampus: Timed Action and Attention

Figure 10 summarizes a model of how the cerebellum adaptively times the release of motor commands by using a "spectrum" of learning sites that are each sensitive to a different range of delays between CS and US. A process of "spectrally timed learning" selects that subset of sites whose reaction rates match the expected delays, or interstimulus intervals (ISIs), between the CS and the US. Learning at parallel fiber/Purkinje cell synapses depresses the tonically active output from cerebellar Purkinje cells to cerebellar nuclei. This Long Term Depression (LTD) occurs in

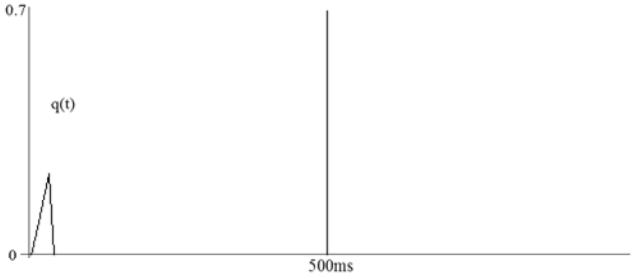
an adaptively timed way. LTD hereby disinhibits the target cerebellar nucleus sites and allows the adaptively timed expression of learned gains from these sites. Such learning enables the cerebellar output to be released at around the time when the US is expected.

In particular, suppose that a conditioned stimulus (CS), say via the motor output pathway *M* in Figure 9, activates pathways to both a subcortical cerebellar nucleus and to cerebellar cortex parallel fibers that synapse on Purkinje cells with a spectrum of differently timed intracellular processes. Unconditioned stimulus (US)-activated climbing fibers provide a teaching signal that also converges upon the parallel fiber/Purkinje cell synapses. This teaching signal causes the active synapses within the parallel fiber spectrum to become weaker (Long Term Depression) if they are activated by the CS when the US teaching signal becomes active. Synapses whose spectral activity does not overlap the climbing fiber signals become stronger (Long Term Potentiation, or LTP). Because the Purkinje cells tonically inhibit their subcortical target cells, their adaptively timed inhibition by the CS disinhibits the effect of tonic Purkinje cell outputs on cerebellar nuclear cells. In other words, a timed gate opens and allows the subcortical cells to fire. The model proposes that climbing fibers also control learning of adaptive gains along subcortical pathways through the nuclear cells. Thus, when the adaptively timed Purkinje cell gate opens, the learned gains can be expressed at the correct time and with the correct amplitude to cause a correctly calibrated motor response.

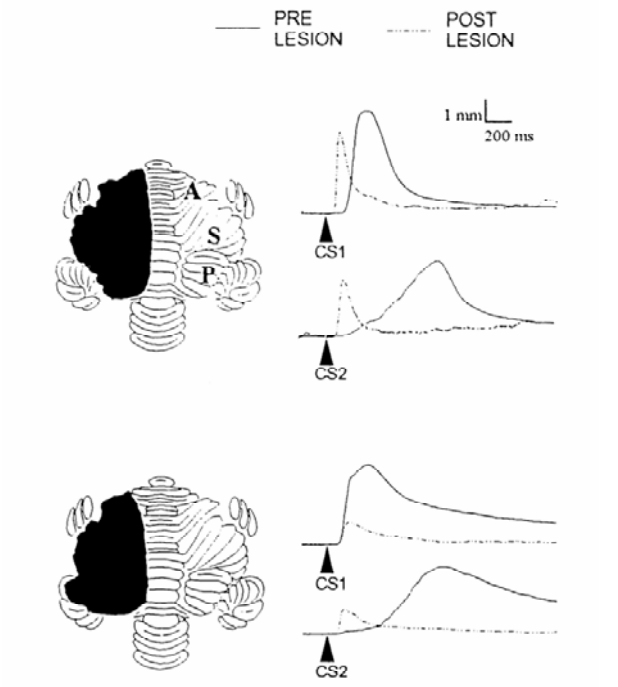
Bullock, Fiala, and Grossberg (1994) and Fiala, Grossberg, and Bullock (1996) have developed and simulated a detailed model of cerebellar adaptive timing. The cerebellar Spectral Timing model of Fiala, Grossberg, and Bullock (1996) links biochemistry, neurophysiology, neuroanatomy, and behavior. This model predicts how the metabotropic glutamate (mGluR) receptor system may be involved in cerebellar adaptively timed learning, as well as, by extension, in other brain regions with adaptively timed cell responses, such as the hippocampus and basal ganglia (Brown, Bullock, and Grossberg, 1999; Grossberg and Merrill, 1992, 1996). A number of subsequent experiments have supported the prediction of a role for calcium signalling and mGluR in cerebellar adaptive timing (Finch and Augustine, 1998; Ichise *et al*., 2000; Miyata *et al.,* 2000; Takechi, Eilers, and Konnerth, 1998). Figure 11a shows a simulated adaptively timed cerebellar output where the interstimulus interval, or ISI, between a conditioned stimulus (CS) and unconditioned stimulus (US) is 500 msec. The series of increasing curves shows the adaptively timed response on successive learning trials. Figure 11b shows how the cerebellar output can be released prematurely when adaptive timing by the cerebellar cortex fails. Figure 11c shows data of Perrett, Ruiz, and Mauk (1993) demonstrating that a similar failure in adaptively timed responding occurs when the cerebellar cortex is lesioned after learning via the nictitating membrane response paradigm. In summary, adaptive timing in the cerebellum reconciles two equally important, but potentially conflicting, behavioral requirements: Fast allocation of attention to motivationally salient events, and adaptively timed responses to these events.



(a)



(b)



(c)

Figure 11. (a) Computer simulation of spectral timing in the cerebellar cortex at the parallel fiber-Purkinje cell synapse. The adaptively timed cerebellar nucleus output after learning in response to an interstimulus interval of 500 msec between the conditioned reinforcer and the unconditioned reinforcer. The increasing curves represent successive learning trials. (b) When the cerebellar cortex is ablated from the model, the response from the cerebellar nucleus occurs prematurely. [Reprinted with permission from Bullock, Fiala, and Grossberg (1994).] (c) Data from the nictitating membrane response paradigm showing a similar effect of a cerebellar cortical lesion on conditioned responses. CS1 was trained to an interstimulus intervals (ISI) or 150 msec, wherease CS2 was trained to an ISI of 750 msec. In both cases, ablation of cerebellar cortex caused a premature response to be made. [Reprinted with permission from Perrett, Ruiz, and Mauk (1993).]

Section 3 summarized data showing that individuals with autism, who are known to have cerebellar deficiencies, also perform short-latency responses in the eye-blink paradigm. The Spectral Timing model provides a way to understand how such adaptive timing deficits occur. Its prediction of a key role of mGluR in adaptively timed learning also points to new experiments or tests that can be done with autistic individuals to determine if, indeed, the mGluR system is not functioning normally in them.

In addition to adaptive timing of motor responses by the cerebellum, there is a need for motivated attention to be maintained long enough for the adaptively timed action to be executed. Adaptively timed motivated attention can hereby prevent irrelevant novel events from prematurely resetting thalamocortical sensory and cognitive representations as they actively read-out adaptively timed responses. The START model (Grossberg and Merrill, 1992, 1996) accomplishes this by showing how circuits within the hippocampus that are capable of adaptively timed learning can modulate the responses of ART and CogEM circuits that have already been summarized. Hoehler and Thompson (1980) have provided experimental evidence that adaptively timed circuits exist in both the hippocampus and the cerebellum by doing ISI shift experiments during which the peak time of the hippocampal trace can change before the peak time of the discrete adaptive response.

In particular, as summarized in Section 5, the ART model proposes how attentional and orienting systems interact to categorize information and to develop resonant states if an active top-down prototype and a bottom-up sensory input form a sufficiently good match. If the mismatch is too big for resonance to occur, then other things being equal, the orienting system can trigger a search for a better category with which to categorize the information. The hippocampal system is proposed to be part of the orienting system that is activated by these mismatches and relays them as novelty-sensitive reset bursts to the thalamocortical system. Such an ART-mediated activation of the orienting system is not, however, sensitive to whether the novel event that caused the mismatch is relevant to the task. The START model clarifies how mismatches may be modulated by task-relevance in an adaptively timed way. In particular, the START model suggests how motivationally salient cognitive representations may be enhanced, while orienting responses are inhibited, by an adaptively timed hippocampal dentate-CA3 circuit, during the same time intervals when conditioned responses are disinhibited by an adaptively timed cerebellar circuit. See Appendix B for a mathematical description of the START Spectral Timing model.

In particular, Figure 9 summarizes how adaptively timed learning within the dentate-CA3 circuits (T in Figure 9) of the hippocampus is proposed to inhibit the activation of the orienting system *A* during an interval wherein a valued and predictable goal is being acted upon. Indeed, hippocampal dentate-CA3 cell firing reflects the learned delays observed during the rabbit nictitating membrane response (Berger, Berry, and Thompson, 1986). Figure 12 summarizes a computer simulation of adaptively timed learning within the dentate-CA3 circuit. The curves in (a) and (b) show how the individual spectral components cooperate to generate an adaptively timed population output *R* in (c). The START model hereby proposes how adaptively timed inhibition of the hippocampal orienting system (Figure 9) and adaptively timed disinhibition of cerebellar nuclear cells (Figure 10) are coordinated to enable motivated attention to be maintained on a goal while adaptively timed responses are released to obtain the goal.

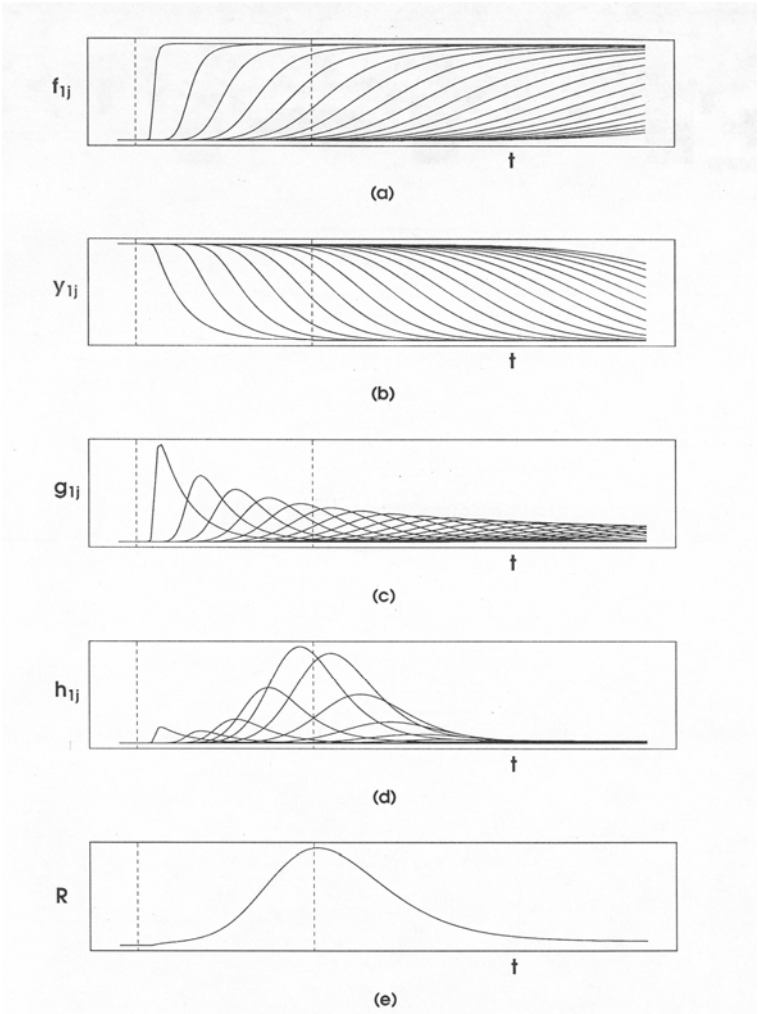


Figure 12. Computer simulation of spectral timing in the dentate-CA3 circuit. (a) The CS_1 input turns on at the time marked by the leftmost vertical dashed line. The US turns on at the time marked by the rightmost vertical dashed line. The functions f_{1j} summarize the different rates of activation of distinct cell sites to the CS_1 input. (b) The functions y_{1j} are chemical transmitters that habituate, or are inactivated, at a rate proportional to their driving signals f_{1j} . (c) The sampling functions g_{1j} are the products of f_{1j} and y_{1j} because the transmitters multiply, or gate, their respective cell activation. The functions g_{1j} are the differently timed responses of cell sites that together form the basis for spectral timing. (d) Learning of the correlation between CS and US occurs at each site only when its g_{1j} is positive. In this sense, each g_{1j} learns by sampling, or gating, the US activity that is correlated with it. Both the timing and rate of learning by the adaptive weights w_{1j} covary with the size of its g_{1j} . Due to the fact that the various g_{1j} have their peak activities at different times, each site is maximally sensitive to learning correlations with different delays between CS and US. The timed cell responses g_{1j} also give rise to outputs h_{1j} in which the signals g_{1j} are multiplied, or gated, by their adaptive weights w_{1j}. In other words, each output $h_{1j} = g_{1j}\, w_{1j}$. (e) When these adaptively weighted signals are added up, they form a total population output signal $R = \sum_j h_{1j}$ that is adaptively timed to peak at around the ISI where the US turns on. Thus, spectral timing is a property of an entire population of adaptively gated pathways. [Reprinted with permission from Grossberg and Merrill (1992).]

Processing stages $S^{(1)}$ and $S^{(2)}$ in Figure 9 play the role of sensory cortex and prefrontal cortex, respectively, in the CogEM model circuit of Figures 4b and 5. Stage D is an emotional center, or drive representation, like the amygdala in Figure 5. Stage M schematizes motor output pathways. The feedback pathways $D \rightarrow S^{(2)} \rightarrow S^{(1)}$ from a particular drive representation to sensory representations are capable of focusing attention on motivationally consistent events in the

world. The excitatory pathways from $S^{(1)} \rightarrow D$ learn the conditioned reinforcer properties of a sensory cue, such as a CS, whereas the pathways $D \rightarrow S^{(2)}$ learn the incentive motivational properties of cues. Representations in $S^{(2)}$ can fire vigorously only if they receive convergent signals from $S^{(1)}$ and D, corresponding to the sensitivity of orbitofrontal cortex to both sensory and reinforcing properties of cues. Then they deliver positive feedback to $S^{(1)}$ and bias the competition among sensory representations to focus attention on their respective features and to attentionally block inhibited features.

Prior to conditioning, a CS can be stored at $S^{(1)}$ and can prime D and $S^{(2)}$ without supraliminally firing these representations, as also illustrated in Figures 4b and 5a. After conditioning, the CS can trigger strong conditioned $S^{(1)} \rightarrow D \rightarrow S^{(2)} \rightarrow S^{(1)}$ feedback and rapidly draw attention to itself as it activates the emotional representations and motivational pathways controlled by D. Representation D can also inhibit the orienting system A as it focuses attention upon motivationally valued sensory events. Here is thus one way in which the CogEm and ART models interact: Emotionally salient goal objects can inhibit the orienting system and thus prevent irrelevant distractors from attracting attention when there is an ART mismatch. This inhibition of the orienting system becomes adaptively timed as follows: The sensory representations $S^{(1)}$ send pathways to a spectral timing circuit T, assumed to be in the dentate-CA3 region of the hippocampus, whose adaptive weights w are trained by a Now Print, or teaching signal, N. The teaching signal N is transiently activated by changes in the activity of the drive representation D that occur when a reinforcing event activates D. After conditioning of T takes place, adaptively timed readout from T can maintain attention on task-relevant cues by amplifying their cortical representations $S^{(2)}$ while inhibiting the orienting system A for an adaptively timed duration. In the figure, the simplest such inhibitory path is depicted, directly from T to D and thereupon to A. A more complex set of pathways exists *in vivo*.

Many data have been rationalized using these circuits, including data from delayed non-match to sample (DNMS) experiments wherein both temporal delays and novelty-sensitive recognition processes are involved (Gaffan, 1974; Mishkin and Delacour, 1975). In summary, as shown in Figures 9 and 10, the START model enables three key properties to simultaneously obtain:

1. *Fast Motivated Attention*. Rapid focusing of attention on motivationally salient cues occurs from regions like the amygdala to prefrontal cortex (the $D \rightarrow S^{(2)}$ pathway in Figure 9). Without further processing, fast activation of the CS-activated $S^{(2)}$ sensory representations could prematurely release motor behaviors.

2. *Adaptively Timed Responding.* Adaptively timed read-out of responses via cerebellar circuits, as in Figure 10, enables learned responses to be released at task-appropriate times, despite the fact that CS cortical representations can be quickly activated by fast motivated attention.

3. *Adaptively Timed Duration of Motivated Attention and Inhibition of Orienting Responses.* Adaptively timed inhibition of mismatch-sensitive cells in the orienting system of the hippocampus (pathway $T \rightarrow D \rightarrow A$ in Figure 9) prevents the premature reset of active CS representations by potentially distracting irrelevant cues during variable task-specific delays. This inhibition is part of the competition that exists between consummatory and orienting behaviors (Staddon, 1983). Even while this inhibitory mechanism prevents CS representations from being prematurely reset, adaptively timed incentive motivational feedback ($D \rightarrow S^{(2)} \rightarrow S^{(1)}$ in Figure 9) helps to maintain the activation of these representations in short-term memory. As a result, the CS representations can continue to read-out the sensory signals that will elicit

adaptively-timed responding. A neural marker of adaptively timed motivational feedback is predicted to be the Contingent Negative Variation, or CNV, event-related potential; see Grossberg (1984a) for further discussion and classical data about CNV.

8. iSTART Model of Autism

8.1. Combining Three Sorts of Imbalances in the iSTART Model

Now that the necessary background models are understood, it is possible to explore how various early-onset lesions influence model properties, and to compare these properties with those of individuals with autism. The main proposal of the present article is that when various START-like mechanisms become imbalanced in the brain — notably, underaroused depression in the drive representations of regions like the amygdala, hypervigilant learning in the recognition learning circuits of temporal and prefrontal cortices, and a failure of adaptive timing in the hippocampal and cerebellar circuits — then formal analogs of behavioral symptoms observed in autism emerge. That is why we call the model the Imbalanced START, or iSTART, model.

8.2. Interactions between Hypervigilant Learning and Underaroused Depression

Sections 4–7 have already summarized some of the behavioral difficulties that these imbalances can cause and, along with Sections 2-3, have compared them with the behavior of individuals with autism. For example, the iSTART model clarifies what happens when underaroused depression is combined with hyperspecific and hypervigilant learning. That is, it explains what happens when an underaroused CogEM model interacts with a hypervigilant ART model. In particular, suppose that positive affect motivates an action. As noted in Section 6.5 and proved in the Appendix, underaroused emotional and sensory dipoles can exhibit a paradoxical enhancement of their ON channels when a nonspecific arousal burst is triggered by an unexpected event. Perseverature bursts of positive affect can hereby be expected from the high probability that the same object, viewed in slightly different ways, will mismatch the hyperspecific top-down expectations of an individual with autism. A persistent perseverative behavior can result, which might manifest itself in persistently inspecting the same object. Suppose, however, that an arousal burst is caused which is significantly larger, say due to a larger mismatch of a presently active hyperspecific prototype with a different and unexpected event in the world. Then an unusually intense, and negative, antagonistic rebound can be caused. Thus, novel experiences can be highly aversive when hyperspecific categories mismatch them and suddenly generate a burst of arousal to the system's underaroused dipoles. These negative rebounds may be one reason why individuals with autism are prone to negative reactions to unanticipated events. One behavioral strategy for coping with these system properties is to avoid the type of novelty that will cause unbearable negative rebounds. The other side of the coin is perseveration on small details of the environment. Such a combination of properties may help to understand the autistic "need for sameness" in many situations, and the fact that behavioral decompensation often occurs in response to what a normal person would view as relatively minor variations in routines.

8.3. Interactions between Underaroused Depression and Adaptive Timing

Let us now consider how adding adaptive timing to the discussion can further clarify how symptoms of autism may arise. In particular, early in development, emotional needs may begin to be met by responding with simple motor patterns in response to basic sensory stimuli. However, continued successful development requires the ability to learn to adaptively time new actions to receive the potentially rewarding consequences of these actions.

To understand this issue better, suppose that new adaptively timed movements cannot be learned. This deficiency can be due to a direct malfunction of adaptive timing circuits in the

cerebellum and/or hippocampus. In particular, if the cerebellar cortex is damaged, then responses can be prematurely released, as illustrated in Figures 11b and 11c. If the dentate-CA3 hippocampal circuit is damaged, then adaptively timed persistence of motivated attention and inhibition of orienting responses cannot occur. If motivated attention and action are not adaptively timed, then spurious resets of attention may more readily occur, as noted in item (3) in Section 7.3. Social skills and language development, in particular, are learned through adaptively timed behaviors in a process of shared attention and imitation. Under these circumstances, a wide variety of social behaviors may not get a chance to be learned, and attention may instead be maintained on lower-order sensory representations and tasks. The above considerations suggest that part of the reason that representations of typical social significance may fail to develop is that they may not get a chance to be strongly reinforced.

Flexible shifts of attention can be impaired because, if the timing circuit T in the hippocampus is damaged (see Figure 9), attention may more easily be distracted from goal objects during task-related delays. Such a lack of timed control over variable delays can harm behavior more when it is necessary to shift attention among different sets of cues. On the other hand, if the orienting system A in the hippocampus is also damaged (see Figures 2 and 9), then flexible reset of attention in response to novel events may be impaired because mismatch-based novelty-mediated attention shifts are no longer operational. If the attentional system in the neocortex remains intact, then direct activation of a recognition code in response to a familiar event is still possible, and the matching process can partially update short-term memory. However, without an intact adaptive timing or orienting system, the network can no longer flexibly search for the proper configuration of targets to attend, especially in the presence of complex spatial layouts that include distracting cues. Gaffan (1992) has described analogous data from hippocampectomized monkeys.

A failure of adaptive timing can also be due to insufficient teaching signals from depressed emotional centers to adaptive timing circuits. Thus, underaroused emotional depression, in addition to its negative effects on the development of a Theory of Mind, can lead to a reduction of normal reinforcing signals to hippocampal and cerebellar circuits, and with it abnormalities of adaptively-timed, motivated attention and action. For example, rewards and punishments generate teaching signals to the adaptive timing circuits in the hippocampus (via $D \rightarrow N$ in Figure 9) and cerebellum (via US-activated climbing fibers in Figure 10) to assure that properly timed responses are reinforced. Reduced Now Print teaching signals N in Figure 9 or climbing fiber teaching signals in Figure 10 can both cause a failure of adaptively timed learning; in the former case, motor learning, and in the later case, motivated attention and inhibition of orienting responses.

If any of these mechanisms fail, then adaptively timed behavior may fail, and with it the future rewards that would normally be received, contingent upon making these adaptively timed behaviorss, may not be forthcoming.

One can also speculate that the breakdown of the normal cycle of behavior and reward, with a dramatic reduction in the normal frequency of behaviorally appropriate rewards, can in itself contribute to a reduction in the arousal I of emotional centers, as in Figure 8 and the Appendix, thereby leading to the types of symptoms, reviewed above, that occur when drive representations D are underaroused. Thus, a tendency towards emotional underarousal can lead to reduced learning signals for adaptively timed learning, which can then lead to a reduction of reward frequency, which can then reinforce the underarousal that prevents adequate reward-based learning signals from occurring. Such a feedback cycle would involve amygdala,

cerebellum, and hippocampus, among other brain regions. In summary, circuits for adaptive timing, reward, motivation, and cognition all interact in the iSTART model via feedback.

8.4. Interactions between Adaptive Timing and Hyperspecific Learning

As another example of how system-wide feedback can maintain, and even worsen, behavioral symptoms, consider the fact that the frequent spurious orienting resets that can occur due to dysfunctional adaptive timing may also contribute to hyperspecific learning. If sensory inputs are prematurely reset, then this can interfere with the normal cycle of adaptively timed shifting of attention to the expected consequences of motor actions. Such a learner could not easily test whether variations on a sensory event predict similar consequences, so abstract prototype formation may not have a chance to occur. This hypothesis predicts that there should be a correlation between the specificity of learned categories in (say) the inferotemporal cortex (Spitzer *et al.*, 1988) and the number of novelty-triggered N200 potentials (Deadwyler *et al.,* 1979, Deadwyler *et al.*, 1981) or even novelty-sensitive cells (Otto and Eichenbaum, 1992) in the hippocampus (Carpenter and Grossberg, 1993).

Drive satisfaction is often contingent on behaviors that use learned abstract prototypes which are capable of recognizing variations of previously experienced events. When drive satisfaction is chronically prevented from occurring as a result of hyperspecific learning, then the drive representations may become depressed. Thus, either hyperspecific learning in neocortical circuits, or failures in adaptive timing in cerebellar or hippocampal circuits, can contribute indirectly to underaroused emotional depression.

A "vicious circle" of environmentally mediated feedback can result in which depressed drives, say as measured by a hypoactive amygdala, fail to trigger the learning of adaptively timed behaviors in hippocampus and cerebellum, whose absence enables the orienting system in hippocampus to spuriously reset cognitive representations in the neocortex during times when attention should be given to a particular task, which then leads to hyperspecific learning of neocortical recognition categories, which then makes it easier to generate mismatch events with sensory cues, which then prevents the normal frequency of behaviorally-appropriate rewards from being received from the amygdala and other reward centers, which then contributes to the maintenance of depressed drives at these centers. Figure 13 summarizes some of these environmentally mediated feedback relationships.

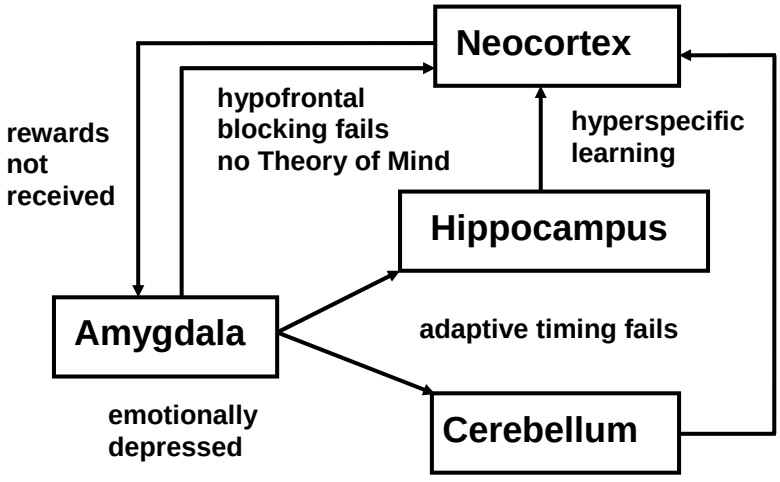


Figure 13. Schematic of feedback relationships whereby particular brain imbalances can ramify throughout other brain regions, thereby sustaining the imbalance by feedback, and also creating or maintaining other symptoms.

In addition to these various environmentally mediated interactions, depressed drive representations may cause a hypofrontal syndrome; see Figure 5b. As a result, the normal motivationally-selective top-down attentional priming signals to sensory cortices will be weakened, attentional blocking will be deficient, and motivationally irrelevant information can flood the sensory system, thereby making it even harder to process motivationally-relevant sensory cues, so that drives continue to be unmet, rewards unreceived, and the cycle perpetuates itself through this route as well.

The motivationally irrelevant information that can flood the brain in this way includes signals from lower-order sensory representations that have built-in pathways to emotional centers. Such a flood of signals can overcome the elevated thresholds due to underarousal of these centers. Excessive emotional responses can result. This property may also clarify the hypersensitivity of individuals with autism to a variety of lower order stimuli, such as noise and touch.

Deficient development of language is clarified by these model mechanisms. Language development requires several factors: It requires shared attention with a caretaker and splitting attention between the objects of that shared attention, analysis of the sounds being produced, of the sounds just heard, and of the motor actions required to make those sounds. The previous sections show how these properties can be damaged in the brain of an individual with autism. In order for language to develop and be learned in a stable way, top-down representations of linguistic structure and word meaning need to be learned (Section 5). Indeed, ART dynamics have been used to explain a variety of data about speech perception and word recognition, including the role of top-down expectations and attention in dynamically grouping evolving sequences of incoming sounds into conscious word percepts (e.g., Grossberg, 2003b; Grossberg, Boardman, and Cohen, 1997; Grossberg and Myers, 2000). Hyperspecific learning can impair the ability of these top-down expectations to incorporate contextually-important language distinctions.

Language development also requires adaptively timed learning to enable children to learn culturally important sequential motor actions through imitation. For example, cerebellar adaptive timing mechanisms, in concert with neocortical working memory mechanisms, have been used to model how children may learn handwriting through imitation (Grossberg and Paine, 2000; Paine, Grossberg, and Van Gemmert, 2005). In the adult, conversation requires sustaining attention and the ability to flexibly disengage it. It requires delaying a motor response appropriately so that reciprocal communication may occur. It requires recognition of representations of social value. All of these processes are impaired within the iSTART model herein described.

The early onset of such imbalances would be expected to disrupt and distort subsequent normal activity-dependent development (Belmonte *et al.* 2004; Herbert, 2004), thereby impeding attention to salient social and communicative stimuli. In comparison, other processing streams may be relatively less affected by dysfunctional adaptive timing and may develop with normal or even enhanced levels of function, due to a lack of competition from hypoactive systems during developmental critical periods. An analogy to visual cortical development as a model may be apt. Visual experience shapes the development of the primary visual cortex using a combination of competitive and associative mechanisms (Grossberg, 1976a; Hubel and Wiesel, 1965; Olson and Grossberg, 1998; von der Malsburg, 1973). Unequal levels of activity result in a greater allocation of cortical resources to the more active eye and a commensurately lesser allocation to the less active eye. The role of activity-dependent mechanisms to shape development has been reported in other sensory cortices (Fox *et al*., 1996; O'Leary *et al*., 1995, Schlaggar *et al*., 1993).

Similar processes may also occur during the development of functional maps that represent higher cortical processes. For example, if linguistic stimuli are subject to adaptive timing deficiencies in children with autism, while musical chord stimuli are not, as was suggested in ERP studies (Dawson *et al*., 1988), then higher cortical resources might be selectively reduced for linguistic processing while they are enhanced for musical processing. Thus autistic development is distorted more than delayed.

Multiple lesions within the cognitive-emotional-timing circuit that is summarized above may combine to result in symptoms of autism. The iSTART model clarifies how several different combinations of deficits can all contribute to a full set of symptoms, consistent with the behavioral, cytoarchitectural, neurophysiological, and genetic data that were summarized in Sections 2 and 3, which together suggest that multiple "hits" may occur in different portions of the brain for autism to fully manifest itself.

8.5. Predictions and Open Issues

The iSTART model leads to a number of testable predictions (see Table 2), and also raises several basic issues that have not previously been articulated, let alone mechanistically explained, and for which experimental evidence does not yet seem to be available. In particular, all of the above explanations of how underaroused depression, hypervigilant learning, and deficient adaptive timing can yield symptoms of autism constitute mechanistic predictions that can be tested. The fact that multiple brain loci support these properties and may all contribute to autistic symptoms is consistent with the existence of significant heterogeneity in the presentation of autism.

New issues include the following: Can underaroused emotional depression and hypervigilant cognitive learning both sometimes be directly caused by a similar underlying defect? This is a reasonable question to ask, because both underaroused depression and hypervigilant learning are problems due to incorrectly calibrated gains: in the case of underaroused depression, the gain of the excitatory signals that arouse the drive representations, say in the amygdala; in the case of hypervigilant learning, the gain of the excitatory signals that try to activate the orienting system, say in the hippocampal system. Alternatively, can one defect indirectly cause the other, as in the case where underaroused depression can weaken adaptive timing, which can lead to spurious "hypervigilant" resets of attention and learning even if the vigilance parameter is chosen in the normal range? Or are they both indirect consequences of a failure within the hippocampal and/or cerebellar adaptive timing circuits themselves? Whichever routes may be there at the outset, the above discussion clarifies how they may perpetuate themselves via a system-wide "vicious circle" of environmentally mediated feedback. The hypothesis that such environmentally mediated feedback can perpetuate or aggravate symptoms also raises the question of how it can be modified to ameliorate them.

Table 2
Model Predictions
Hyperspecific learning of recognition categories

Vary recognition exemplars gradually to test for narrow category boundaries

Test ERPs that measure category reset, memory search, and attention shifts; e.g., correlated series of P120, N200, and P300 potentials

Test fMRI measures of abnormal novelty detection in hippocampal area

Test the predicted relationship between increased specificity of learned categories, say in the temporal and prefrontal cortices, and the number of novelty-triggered N200 ERPs

Merge the Spitzer et al (1988) and Otto and Eichenbaum (1992) neurophysiological paradigms to test in monkeys if difficult discrimination conditions generate more hippocampal novelty potentials

Study how baseline vigilance level is determined in normal individuals and clinical patients; cf. a possible role for Acetylcholine

Test if some individuals with autism have abnormally high baseline vigilance while some individuals with medial temporal amnesia may have abnormally low baseline vigilance

Study how baseline vigilance may change during memory search for a correct recognition category; correlate vigilance changes with N200 and P300 ERPs

Test if the mismatch of a more general category causes vigilance to momentarily increase more than mismatch of a more specific category ("match tracking")

Test if hyperspecific learned categories prevent unrecognized variations of an event from becoming conditioned reinforcers due to their inability to trigger the actions that elicit the normal number of rewards

Underaroused Emotional Depression

Test if certain behavioral thresholds are elevated but responses are hypersensitive when the elevated threshold is exceeded by carefully controlled increments in stimulus intensity

Study possible brain bases for underaroused depression, notably in the amygdala

Use classical conditioning blocking paradigm to test if attentional blocking is deficient

Correlate blocking properties with fMRI measures of amygdala unresponsiveness; does poor blocking correlate with amygdala unresponsiveness?

Test if orbitofrontal cortex is less responsive during intervals of amygdala unresponsiveness

Test if low amygdala activation occurs during inadequate attentional blocking while high amygdala activation occurs during events that cause extreme negative emotional reactivity; e.g., variations in routines and responses to some sensory stimuli

Test if small mismatches of a hyperspecific recognition category do not generate a P300 ERP but large mismatches trigger both a P300 and strong amygdala activation corresponding to negative affect

Study ERPS and fMRI measures during perseverative attention to small details of an object for evidence that enhanced positive motivational activations, including ON motivational bursts, may occur during slight shifts of attention, rather than the extreme negative emotional activations that larger variations in routine and some sensory stimuli may cause

Develop tests to see if reducing the proposed underaroused depression by gradually increasing arousal results in less severe behavioral reactions to immediately subsequent suprathreshold stimuli

Failures of Adaptively Timed Learning

Test if adaptive timing deficits occur in certain autistic individuals, whether due to premature release of actions as a result of a cerebellar problem, or inability to maintain attention upon motivationally valued goal objects due to a problem in the dentate-CA3 hippocampal area

Study possible cerebellar deficiencies using the eye blink paradigm, including possible problems with metabotropic glutamate receptor functioning, for evidence of problems with adaptive timing of actions.

Study possible hippocampal deficiencies by testing inability to maintain attention upon a valued goal object for a predetermined time interval

Test if some failures of adaptive timing may be due to inadequate Now Print teaching signals in hippocampal and cerebellar circuits due to inadequate responses in areas like the amygdala to rewarding events

Use adaptive timing deficits as one marker for future autism in some infants

By clarifying that several different routes to autism may exist, the iSTART model also opens the way towards differentiating individuals with autism early in life in terms of which route they may be on. Such a clinical differentiation may lead to more effective treatments in the future.

9. Discussion

9.1. Comparison with other Theories of Autism

In Section 3. the requirements of a satisfactory model of autism were set forth. The iSTART model provides a conceptual and mechanistic foundation that substantially meets these requirements. As explained above, iSTART explains how the phenotypic common final pathway of autism can result from a variety of different combinations of early onset imbalanced perturbations to the system and is thus consistent with both the clinical and the etiologic heterogeneity of autistic spectrum disorders. Indeed, it has already been suggested by others that clinical, imaging and structural heterogeneity of autistic spectrum disorders may best be explained in terms of disruption of a distributed circuit that includes cerebellar, limbic, and cortical structures, and not necessarily to any particular point of the circuit (Eigsti and Shapiro, 2003; Herbert, 2004; Lee *et al,*. 2003).

The iSTART model is consistent with a large experimental database and proposes how key autistic features arise. It clarifies how social functions and attention, verbal and nonverbal communication, and imitation skills may be particular affected, how a concrete hyperspecific cognitive style results, and how development subsequently distorts. It provides an explanation of the apparent paradoxical co-occurance of hyporesponsivity to many, particularly social, stimuli, with extreme negative reactivity to many lower-order stimuli, changes in routine, and social stimuli that manage to overcome an elevated threshold. It is consistent with the frequent presence of motor dysfunction and dyscoordination in autistic individuals and its apparent early onset, and with the documented deficiencies of autistic individuals in the adaptive timing task of the conditioned eye blink response and in motor planning. In this last regard, the model is also notable for being consistent with the presence of subtle deficiencies in early development, escalating into more florid developmental distortions as higher-order processing demands increase. It provides a more precise analysis of how breakdowns of brain mechanisms can lead to symptoms of autism than other proposed models of autism. While it is compatible with many of these alternative models, it also differs from them by explaining how each of their hypotheses may be manifestations of specific breakdowns in prescribed brain mechanisms, in anatomically specified regions, interacting within a dysfunctional network with other anatomically specified brain mechanisms, rather than being prime causes.

One alternative model (Howard *et al*., 2000) supposes autism to be secondary to an amygdala deficit. This hypothesis is based on studies showing that the amygdala of individuals with autism fails to activate in normal ways to a variety of social stimuli such as faces. There are also similarities in the neuropsychological profiles of high functioning individuals with autism and other patients with amygdala damage. The iSTART model is compatible with the amygdala hypothesis because it considers a depressed, notably underaroused depressed, drive representation within a region like the amygdala to be a critical network feature. The iSTART model also notes, however, that symptoms of autism may arise even if the prime lesion is not in the amygdala, since amygdala depression can also result from imbalances elsewhere in the brain.

A similar comment can be made about the executive dysfunction hypothesis (Hughs *et al*., 1996). Here again, the iSTART model includes executive dysfunction *as a result of* underaroused depression. While hypofrontality, and thus executive dysfunction, could in principle cause depression in drive circuits by using their reciprocal connections, there does not

seem to be experimental evidence for the prime lesion of autism as occurring in the frontal lobes. Executive dysfunction is also not the earliest manifestation of the condition. For example, a study of preschoolers with autism found no group differences between them and normal controls on eight executive function tasks, but did find that children with autism initiated fewer joint attention and social interaction behaviors (Griffen *et al*., 1999). This result has been confirmed in a larger study, comparing children with autistic spectrum disorder, children with developmental delay, and normal children, matched for mental age, on both dorsolateral and ventromedial prefrontal tasks (Dawson *et al*., 2002). Children with autistic spectrum disorders performed comparably to both comparison groups on all executive function tasks. Moreover, executive function deficits are seen in other disorders (Volkmar and Pauls, 2003) and do not correlate well with degree of social disability (Dawson *et al.* 1998).

A deficient Theory of Mind (TOM) has also been proposed to explain autism (Baron-Cohen, 1989; Perner *et al*., 1989). In this proposal, autism is the result of an inability to understand the mental states and motivations of others. Consequently they are unable to competently predict others' actions and therefore perform actions that are socially aberrant. This proposal well describes many (but not all) aspects of older childhood and adult autistic behavior, but is subject to the same criticisms as is the executive function hypothesis. First of all, autism occasionally becomes manifest before TOM is felt to be extant: The earliest estimate for operational TOM is about 18 months and explicit operation of TOM is not felt to occur before age 4 to 6 years (Frith and Frith, 2003), whereas children with autism have measurable problems with eye contact, orienting to name, joint attention, imitation, nonverbal communication, and language development before 18 months of age (Charman *et al*., 1997; Cox *et al*., 1999). Indeed, behaviors that distinguish infants with autism from other developmental disabilities are identifiable as early as 8 months of age (Baranek, 1999; Mars *et al*., 1998; Werner *et al*., 2000). Deficient TOM is not specific to autism (Happe and Frith, 1996). Higher functioning autistic individuals have been shown to have performed adequately at TOM tasks while still having severe social dysfunction (Bowler, 1992; Sigman, Yarmiya and Capps. 1995). Also, as for the amygdala and the executive dysfunction hypotheses, deficient TOM may occur as a result of underaroused depression in iSTART.

Another group of related hypotheses are the "weak central coherence" model (Happe, 1996) and the "deficient hierarchization" model (Mottron *et al*., 1999). While differing in subtle ways, both of these hypotheses focus on deficiencies in binding perceptual inputs into higher-order representations. Neither of these models explain all key autistic features, or attempt to fit known neuropathologic data. The iSTART model proposes an ART-based learning mechanism whereby this processing deficiency can occur and places it in context wherein many data about normal learning and binding can be explained.

Several groups have used neural networks to model "weak central coherence" and the cognitive hyperspecificity that is commonly seen in autism. Gustafsson has described autism as deficient self-organization of feature maps (Gustafsson, 1997). Gustafsson initially speculated that excessive lateral inhibition, as a primary deficit, may prevent adequate feature maps from forming, and more recently has proposed a low capacity to produce serotonin and insufficient nitric oxide production as possible mechanisms (Gustafsson 2004). Brock and colleagues (Brock *et al,.* 2002) proposed that a deficit in temporal binding results in a reduction of integration of various different specialized local neural networks and consequently weak central coherence, and O'Laughlin and Thagard (O'Laughlin and Thagard, 2000) modelled weak central coherence as a consequence of a high inhibition to excitation ratio in a connectionist network.

"Weak central coherence" is explained within the iSTART model as a manifestation of hypervigilant ART-based learning. Indeed, ART models contain self-organizing feature maps as part of their dynamics, notably the bottom-up adaptive pathways and competitive selection of recognition categories in Figure 2a. iSTART differs from these other models by demonstrating how hypervigilant learning may result from many different sorts of early-onset imbalanced brain mechanisms and places it within a system that explains much more of autism than just the hyperspecific cognitive style.

McClelland (2000) contrasted cross-bar associative networks, which can cause massive associative interference, with conjunctive codes that can learn compressed categories and thereby reduce the amount of associative interference. He illustrated conjunctive coding using a back propagation model. McClelland proposed that, through some "subtle change in some of the parameters…they may be predisposed to use an excessively conjunctive form of neural coding" (p. 501). This property was suggested to clarify hyperspecificity in autism. What parameters may be involved were not specified. Unfortunately, back propagation is widely understood to be biologically implausible (Grossberg, 1988). The back propagation model does not have any top-down attentive matching, uses non-local transport of learned weights, cannot learn without supervision, can learn only slowly, and has an unstable memory in response to either changing exemplar statistics, fast learning, or even sustained maintenance of an input pattern. In contrast, categorization within an ART model can prevent associative interference at all vigilance levels, while high vigilance can lead to hyperspecific categories which can too easily mismatch input exemplars.

Cohen (1994) also used a back propagation model, but not to simulate possible brain mechanisms of autism. Rather, Cohen used the model to learn a map with a training set formed from interviews with caregivers of children with autism and caregivers of children with mental retardation not associated with autism. The map was learned between eleven descriptive properties of behavior (reactions to pain, eye contract, gesture, imagination, tactile defensiveness, social interaction, verbal perseveration, topic perseveration, repetitive behavior, empathy/facial expression, intonation/understanding) and the two outputs (autism, mental retardation). The test set consisted of interviews of parents of children with autism and of children with mental retardation not associated with autism. From this learned map, Cohen (1994) concluded that "having too few processing elements led to relatively weak learning and generalization while having too many processing elements…led to good learning but relatively weak generalization" (p. 13). These are conclusions, however, about how back propagation can learn which combinations of interview categories predict autism, not about how the brains of individuals with autism work.

Courchesne and Allen have proposed that the parietal lobe and the cerebellum are both involved in the pathophysiology of autism with cerebellar modulation of the use of attentional resources (Allen and Courchesne, 2001). The iSTART model is compatible with this proposal, and shows how deficiencies of basic neocortical ART attentional mechanisms and basic cerebellar adaptive timing mechanisms might contribute to symptoms of autism. Moreover, iSTART also describes how interacting mechanisms within a larger brain system give rise to symptoms of autism that go beyond the Courchesne proposal.

9.2. Concluding Remarks

The Imbalanced START model proposes how particular types of imbalanced mechanisms in different parts of the brain can generate autistic symptoms through brain-wide interactions. The model proposes a unified explanation of why autism is characterized by deficient imitation skills,

deficient social skills, hypervigilant learning, uneven cognitive strengths and weaknesses, and hypersensitivity to various stimuli. Multiple paths can lead to this stable system state: primary early underaroused depression, or primary early dysfunctional adaptive timing, or combinations of each can lead to a "vicious circle" of environmentally mediated feedback. This approach enables the iSTART model to provide a rigorous conceptual framework within which to unify the major previous models of autism, and to clarify how the hypotheses of these models may be mechanistically explained.

Although the iSTART model provides testable linking hypotheses between brain mechanisms and behavioral symptoms, it is not without its limitations. In its present form, the model can explain and predict how various combinations of imbalanced interactions between cognitive, emotional, and timing systems can lead to symptoms of autism, but it does not explain their underlying genetic or biochemical causes. However, by clarifying links between underlying brain mechanisms and behavioral outcomes, the iSTART model may help future research to focus more directly on characterizing the types of mechanisms that can lead to these behavioral outcomes, and how their underlying genetic or biochemical causes may be linked.

To this end, the iSTART model makes testable predictions, several of which are outlined above. Signs of underaroused depression and of dysfunctional adaptive timing should be found early in infants destined to develop autism. A means of quantifying either or both of these functions in infants would be a valuable tool in testing cohorts of high-risk infants. In particular, during underaroused depression, certain emotional centers, such as the amygdala, may be less easily triggered, but may have excessive responses when their elevated thresholds are exceeded, say in response to particular smells. Excessive responses from such emotional centers may be measured by fMRI. Likewise, sudden but partial reductions in an affectively charged cue may cause an affective rebound in normals, much as withdrawal of a fearful cue can cause relief, or removal of ice cream can cause frustration. During underaroused depression, such an antagonistic rebound may not occur. (On the other hand, it can cause other behavioral effects due to hypervigilance.) Additional testable markers of underaroused depression are summarized in the Appendix.

The iSTART model reinforces the importance of early identification and treatment of autistic disorders. Early identification could allow for interventions which disrupt the cycle of feedback that can develop between underaroused depression of drive representations and dysfunctional adaptive timing. Early overt and prolonged pairing of important socio-emotional and language-related stimuli with salient reinforcers may be able to partially compensate for underaroused depression and dysfunctional adaptive timing. Such interventions may help relevant processing streams win the competition for processing resources and thereby mitigate the distortions of development that characterize autism.

GRANT SUPPORT AND OTHER ACKNOWLEDGEMENTS

Stephen Grossberg was supported in part by the National Science Foundation (NSF SBE-0354378) and the Office of Naval Research (ONR N00014–01–1–0624). The authors would like to thank Robin Amos and Susanne Daley for their expert help in the preparation of the manuscript and figures.

APPENDIX A:
HABITUATIVE TRANSMITTER S AND GATED DIPOLE OPPONENT PROCESSES

This appendix demonstrates a number of testable properties of habituative transmitters and gated dipole opponent processing circuits that are relevant to the hypothesis that various such dipoles are underaroused in many individuals with autism.

A1. Habituative Transmitters as Gates at Depressing Synapses. Gated dipole opponent processes rely heavily for their properties on how chemical transmitters transform signal flow through the dipole circuit. Gated dipoles use a chemical transmitter model that was derived from associative learning postulates in Grossberg (1968, 1969c). The gated dipole model itself was derived from conditioning postulates in Grossberg (1972b). The transmitter derivation that is given below argues that this transmitter law is the minimal dynamic law for unbiased transmission using a depletable signal (Grossberg, 1980). This type of law has recently received additional experimental support and has also been called a law for depressing synapses (Abbott et al., 1997; Tsodyks, Pawelzik, and Markram, 1998). The experimental and modeling literature on this topic has grown rapidly during the past several years.

We start by asking the following question: What is the simplest law whereby one nerve cell can send unbiased signals to another nerve cell? The simplest law says that if a signal S passes through a given nerve cell v_1, the signal has a proportional effect

$$T = SB, \tag{A1}$$

where $B > O$, on the next nerve cell v_2. Suppose, in addition, that the signal from v_1 to v_2 is due to the release of a chemical *z(t)* from v_1 that activates v_2. If such a chemical transmitter is persistently released when S is large, what keeps the net signal, T, from getting smaller and smaller as v_1 runs out of transmitter? Some means of replenishing or accumulating the transmitter must exist to counterbalance its depletion due to release from v_1. To accommodate this interpretation, we can rewrite (A1) in the form

$$T = Sz \tag{A2}$$

and ask: How can the system keep z replenished so that

$$z(t) \cong B \tag{A3}$$

at all times *t*? This is a question about the sensitivity of v_2 to signals from v_1, since if z could decrease to small values, then even large signals S would have only a small effect on T. Equation (A2) has the following biophysical interpretation. The signal, S, causes the transmitter, z, to be released at a rate $T = Sz$. When two processes, such as S and z, are multiplied, they are said to interact by *mass action*, or that z *gates* S. In summary, (A2) says that z gates S to release a net signal T, and (3) says that the cell tries to replenish z to maintain the system's sensitivity to S. The simplest law that joins together both equations (A2) and (A3) is the following differential equation for the net rate of change, dz/dt, of z:

$$\frac{dz}{dt} = A(B - z) - Sz\,. \tag{A4}$$

Equation (A4) describes the following four processes going on simultaneously.

Accumulation or Production and Feedback Inhibition: The term $A(B - z)$ can be given two possible interpretations, depending on whether it represents a passive accumulation process or an active production process. In the former case, there exist B sites to which transmitter can be bound, z sites are bound at time *t*, and $B - z$ sites are unbound. Then term $A(B - z)$ says that transmitter is bound at a rate proportional to the number of unbound sites. In the latter interpretation, two processes go on simultaneously. Term AB on the right-hand side of (A4) says

that z is produced at a constant rate AB. Term $-Az$ says that once z is produced, it inhibits the production rate by an amount proportional to the concentration of z. In biochemistry, such an inhibitory effect is called feedback inhibition by the end product of a reaction. Without feedback inhibition, the constant rate of production, AB, would eventually cause the cell to burst. With feedback inhibition, the net production rate is $A(B-z)$, which causes $z(t)$ to approach the finite amount B, as is required by (A3). The term $A(B-z)$ hereby enables the cell to accumulate a target level B of transmitter.

Gating and Release: Term $-Sz$ in (A4) says that z is inactivated or released at a rate Sz. As in (A2), inactivation or release of z is due to a mass action interaction, or gating, of S by z.

Equations (A2) and (A4) describe the simplest dynamical law that corresponds to constraints (A2) and (A3). These equations reconcile the two constraints of unbiased signal transmission and maintenance of sensitivity when the signals are due to release of transmitter.

A2. Weber-Law Adaptation and Habituation. To determine how the net signal, $T = Sz$, reacts to a sudden change in S, as in Figure 8a, suppose that $z(t)$ reacts slowly compared to the rate with which $S(t)$ can change. For definiteness, suppose that $S(t) = S_0$ for all times $t \le t_0$ and that, at time $t = t_0$, $S(t)$ suddenly increases to S_1. In the ON channel of Figure 8a, $S_0 = f(I)$ and $S_1 = f(I + J)$. By (A4), $z(t)$ reacts to the constant value $S(t) = S_0$ by approaching an equilibrium value $z(t_0)$. This equilibrium value is found by setting $dz/dt = 0$ in equation (A4) and solving for

$$z(t_0) = \frac{AB}{A + S_0}. \tag{A5}$$

By equation (A5), a larger value of S_0 causes more transmitter to be inactivated or released. In other words, $z(t_0)$ is a *decreasing* function of S_0. By contrast, (A2) implies that the net signal to v_2 at time t_0 equals

$$S_0 z(t_0) = \frac{ABS_0}{A + S_0}, \tag{A6}$$

and thus that the rate of transmitter release is an *increasing* function of S_0. Thus, even though the transmitter is depleted more by a larger input signal, the net output signal grows with the input.

Now let $S(t)$ switch to the value $S_1 > S_0$. Because $z(t)$ is slowly varying, $z(t)$ approximately equals $z(t_0)$ for awhile after $t = t_0$. Thus, the net signal to v_2 during these times is approximately equal to

$$S_1 z(t_0) = \frac{ABS_1}{A + S_0}. \tag{A7}$$

Equation (A7) has the same form as a Weber law, $K(A + L)^{-1}$. The signal S_1 is evaluated relative to the baseline, S_0, just as K is evaluated relative to L. This Weber law is due to slow intracellular adaptation of the transmitter gate to the input level through time. It is not due to fast intercellular lateral inhibition across space (Grossberg, 1980, Appendix C and D), which also obeys a Weber law, and which provides key normalization properties during categorization within an ART model. Many of the properties derived below are due to this intracellular Weber law.

As $z(t)$ in (A4) begins to respond to the new transmitter level, $S = S_1$, $z(t)$ gradually approaches the new equilibrium point that is determined by $S = S_1$, namely

$$z(\infty) = \frac{AB}{A + S_1}. \tag{A8}$$

The net signal consequently decays to the asymptote,

$$S_1 z(\infty) = \frac{ABS_1}{A + S_1}. \tag{A9}$$

Thus, after $S(t)$ switches from S_0 to S_1, the net signal Sz jumps from (A6) to (A7) and then gradually decays to (A9). See x_3 in Figure 8a. The exact course of this decay is described by the equation

$$S_1 z(t) = \frac{ABS_1}{A + S_0} e^{-(A+S_1)(t-t_0)} + \frac{ABS_1}{A + S_1}(1 - e^{-(A+S_1)(t-t_0)}) \tag{A10}$$

for $t \geq t_0$, which shows that the rate, or gain, $A + S_1$ of the response increases with the signal S_1, just as in the case of shunting lateral inhibition (Grossberg, 1980). The sudden increment followed by slow decay can be intuitively described as an overshoot followed by habituation to the new sustained signal level, S_1. Both intracellular adaptation and habituation occur whenever a transmitter fluctuates more slowly than the signals that it gates.

The size of the overshoot can be found by subtracting equation (A9) from (A7). As in the ON channel of Figure 8a, let $S_0 = f(I)$ and $S_1 = f(I + J)$, where $f(w)$ is a signal function that tranforms the inputs I and $I + J$ that exist before and after the increment J into net signals S_0 and S_1, respectively. Then the overshoot size is approximately

$$S_1 z(t_1) - S_1 z(\infty) = \frac{ABf(I + J)[f(I + J) - f(I)]}{[A + f(I)][A + f(I + J)]}. \tag{A11}$$

Section 4 below shows that the rebound size in response to offset of the phasic input J is related to (A11) in a way that allows both $f(w)$ and the tonic arousal level, I, to be estimated.

A3. Gated Dipole. It is shown below how, if transmitters gate signals before the gated signals compete, as in Figure 8a, then an antagonistic rebound can be elicited by offset of a specific phasic input, as in light-ON *versus* light-OFF, or fear *versus* relief. It is also shown how unexpected events can cause an antagonistic rebound. They do this by triggering a sudden increase in the level of nonspecific arousal that is gated by all the transmitter pathways.

Figure 8a depicts the simplest network in which two opponent ON and OFF channels receive inputs that are gated by slowly varying transmitters before the channels compete to elicit a net output response. In such a feedforward gated dipole, specific phasic inputs are turned on and off by internal or external cues and nonspecific arousal inputs are on all the time, or tonic, even though their size can vary through time. Each channel can have its own sum of specific inputs, K_1 or K_2, such as hunger or satiety drive inputs, respectively, that are added to positive or negative conditioned reinforcer signals. Both channels also receive the same arousal input, L. The total signals to the two channels are, therefore, $S_1 = f(K_1 + L)$, where the signal function, $f(w)$, is monotone increasing.

The relative sizes of S_1 and S_2 and their rates of change through time relative to the transmitter fluctuation rate determine whether an antagonistic rebound will occur. To emphasize this fact, let

$$I = \min(K_1 + L, K_2 + L) \tag{A12}$$

and

$$J = |K_1 - K_2|. \tag{A13}$$

In other words, the smaller slowly varying total input to a dipole channel, I, determines the network's net arousal level and J determines how asymmetric the inputs are to the two channels. Suppose, for definiteness, that $K_1 > K_2$. Then $S_1 = f(I + J)$ and $S_2 = f(I)$. The notational shift from $S_1 = f(K_1 + L)$ and $S_2 = f(K_2 + L)$ to $S_1 = f(I + J)$ and $S_2 = f(I)$ in equations (A12) and (A13) is motivated by more than formal convenience. The notation I and J emphasizes that the dipole does not know how many input sources are perturbing it through time. All it can compute is the net arousal level, I, and the degree of asymmetry, J, above I, whether one or a million input sources are active. If a million cues equally perturb the ON-channel (e.g., positive conditioned reinforcers, as in Figures 4 and 5) and another million cues equally perturb the OFF-channel (e.g., negative conditioned reinforcers), the net effect of all the cues will be to increase I, not J. Thus, after dipole competition takes place, all these cues need not generate any incentive motivation. On the other hand, by increasing I, these cues can alter the sensitivity of the dipole to other asymmetrically distributed inputs due to the dipole's Inverted-U properties; see Section A6 below. This is the kind of simple but subtle distinction that the I and J notation represents in Figure 8a.

A4. Rebound due to Phasic Cue Offset. One of the diagnostic properties of a gated dipole opponent process is that it generates antagonistic rebounds in response to sudden changes in both its phasic input *J* and its tonic arousal input *I*. How these rebound properties are predicted to change when a dipole receives insufficient arousal ("underaroused depression") is one of the testable properties of the iSTART model; see Sections 6.5 and 6.6. A rebound can be caused, after the network equilibrates to the phasic input J, if input *J* is suddenly shut off. This effect is analogous to the reaction that occurs when a previously sustained light is shut off in a sensory dipole, or a previously sustained aversive cue is shut off in a motivational dipole. To see how this rebound is generated, suppose that the arousal level is I and that the cue input is J. Let the total signal in the ON-channel be $S_1 = f(I + J)$ and that in the OFF-channel be $S_2 = f(I)$; see Figure 8a. Let the transmitter in the ON-channel, z_1, satisfy the equation

$$\frac{d}{dt} z_1 = A(B - z_1) - S_1 z_1 \tag{A14}$$

and the transmitter in the OFF-channel, z_2. satisfy the equation

$$\frac{d}{dt} z_2 = A(B - z_2) - S_2 z_2 . \tag{A15}$$

After z_1 and z_2 equilibrate to S_1 and S_2, their levels no longer change through time, so $(d/dt)z_1 = (d/dt)z_2 = 0$. At equilibrium, equations (A14) and (A15) imply

$$z_1 = \frac{AB}{A + S_1} \tag{A16}$$

and

$$z_2 = \frac{AB}{A + S_2} . \tag{A17}$$

Since $S_1 > S_2$, it follows that $z_1 < z_2$*;* that is, z_1 is habituated more than z_2. However, the gated signal in the ON-channel is $S_1 z_1$ and the gated signal in the OFF-channel is $S_2 z_2$. Since

$$S_1 z_1 = \frac{ABS_1}{A + S_1} \tag{A18}$$

and

$$S_2 z_2 = \frac{ABS_2}{A + S_2}, \tag{A19}$$

it follows from the inequality $S_1 > S_2$ that $S_1 z_1 > S_2 z_2$, despite the fact that $z_1 < z_2$. Thus, the ON-channel gets a bigger signal than the OFF-channel. After the two channels compete, the input J produces a sustained ON-output whose size is proportional to

$$S_1 z_1 - S_2 z_2 = \frac{A^2 B[f(I+J) - f(I)]}{[A + f(I)][A + f(I+J)]}. \tag{A20}$$

Division of the overshoot amplitude (A11) by the sustained ON-output amplitude (A20) yields a predicted relationship between the size of the overshoot in the ON-channel and the size of the steady-state ON-output; namely,

$$\frac{on - overshoot}{steady\ on - output} = \frac{f(I+J)}{A}, \tag{A21}$$

which provides an estimate of $f(w)$ and I if J is parametrically varied.

In particular, if $f(w)$ is a linear signal, $f(w) = w$, then the sustained ON-ouput in (A20) becomes

$$S_1 z_1 - S_2 z_2 = \frac{A^2 BJ}{(A + I)(A + I + J)}, \tag{A22}$$

which is an increasing function of J (e.g., more fear given more shock) but a decreasing function of the arousal level I (analgesic effect).

Now shut J off to see how an antagonistic rebound (e.g., relief) is generated. The cell potentials rapidly adjust until new signal values, $S_1^* = f(I)$ and $S_2^* = f(I)$, obtain. However, the transmitters z_1 and z_2 change much more slowly, so that equations (A16) and (A17) are approximately valid in a time interval that follows J offset. Thus, the gated signals in this time interval approximately equal

$$S_1^* z_1 \cong \frac{ABf(I)}{A + f(I+J)} \tag{A23}$$

and

$$S_2^* z_2 \cong \frac{ABf(I)}{A + f(I)}. \tag{A24}$$

Thus, $S_1^* z_1 < S_2^* z_2$. The OFF-channel now gets the bigger signal, so an antagonistic rebound occurs, the size of which is approximately

$$S_2^* z_2 - S_1^* z_1 = \frac{ABf(I)[f(I+J) - f(I)]}{[A + f(I)][A + f(I+J)]}. \tag{A25}$$

Division of the rebound amplitude (25) by the steady-state ON-output (20) yields an interesting relationship between the maximal OFF-rebound-output and the steady ON-output; namely,

$$\frac{off - rebound}{on - output} = \frac{f(I)}{A}, \tag{A26}$$

which, combined with the estimate of f and I from (A21), provides a sharper estimate of the arousal level I. A comparison of equation (A21) with (A26) shows that, as the arousal level I is parametrically varied, (A21) should have the same graph as (A26), shifted by J. This comparison provides an estimate of J (that is, of how the behavioral input is transformed into

neural units) and also a strong test of the model. Once $f(w)$ is estimated, equations (A20) and (25) can be verified. If the signal f is linear, so that $f(w) = w$ in (A25), then

$$S_2^* z_2 = S_1^* z_1 = \frac{ABIJ}{(A+I)(A+I+J)}. \quad \text{(A27)}$$

The rebound is then an increasing function of J (e.g., offset of a larger shock elicits more relief) and the dipole rebound is an Inverted-U function of I (an optimal arousal level exists).

The rebound is transient (see OFF in Figure 8a) because the equal signals, $S_1 = S_2 = f(I)$ gradually equalize the z_1 and z_2 transmitter levels until they both approach $AB(A+f(I))^{-1}$. Then $S_1 z_1 - S_2 z_2$ approaches zero, so the competition between channels shuts off both of their outputs.

A5. Rebound due to Mismatch-triggered Arousal Burst. A surprising property of gated dipoles is their reaction to sudden increments, or bursts, in the arousal level, I. Such increments may, for example, occur in response to unexpected events. In an ART model, mismatch between a learned top-down expectation and a bottom-up sensory or cognitive input pattern can trigger a burst of nonspecific arousal from the orienting system; see Figure 2c. An arousal-mediated rebound enables ongoing information processing to be reset in response to the predictive disconfirmation that the mismatch represents.

In order to calibrate the predicted size of the rebound, suppose that the ON-channel and the OFF-channel have equilibrated to the input levels I and J. Now suddenly increase the arousal level I to I^*, thereby changing the signals to $S_1^* = f(I^* + J)$ and $S_2^* = f(I^*)$ The transmitters z_1 and z_2 continue to obey equations (A16) and (A17) for awhile, with $S_1 = f(I+J)$ and $S_2 = f(I)$. A rebound occurs if $S_2^* z_2 > S_1^* z_1$. In general,

$$S_2^* z_2 - S_1^* z_1 = \frac{AB[f(I^*) - f(I^*+J)] + B[f(I^*)f(I+J) - f(I)f(I^*+J)]}{[A+f(I)][A+f(I+J)]}. \quad \text{(A28)}$$

In particular, if the signal function is linear, with $f(w) = w$, then a rebound occurs whenever

$$I^* > I + A, \quad \text{(A29)}$$

since then

$$S_2^* z_2 - S_1^* z_1 = \frac{ABJ(I^* - I - A)}{(A+I+J)(A+I)}. \quad \text{(A30)}$$

Thus, given a linear signal function, a rebound will occur if I^* exceeds $I + A$ no matter how J is chosen. In other words, if the event is so unexpected that it increments the arousal level by more than amount A, then all dipoles in the network will simultaneously rebound. Moreover, the size of the OFF-cell rebound increases as a function of the size of the ON-cell input, J, as equation (A30) shows. In particular, no rebound occurs if the ON-cell was inactive before the unexpected event occurs. The rebound mechanism is selective: It rebounds most vigorously those cells which are most active ($J >> 0$), and which thus played the largest role in reading out the mismatched expectation, and spares inactive cells ($J \cong 0$).

A6. Inverted-*U* in Dipole Output. The signal function f in a gated dipole is not, in general, linear. In fact, it is often a sigmoid, or S-shaped, function of activity (Grossberg, 1973, 1980). Sigmoid signal functions have important properties of noise suppression and contrast-enhancement. The predicted underaroused depression of an individual with autism follows from an Inverted-U property; see Figure 7. A gated dipole exhibits an Inverted-U property if its signal

function $f(w)$ is a sigmoid function; that is, if $f(0) = df/dw(0) = 0$, $df/dw(w) > 0$ if $w > 0$, $f(\infty) < \infty$, and $d^2 f/dw^2(w)$ changes sign once from positive to negative as w increases. In particular, if $f(w)$ is sigmoid, an Inverted-U occurs in the sustained ON-output (A20) as I is parametrically increased, despite the fact that an Inverted-U does not obtain in (A22) when $f(w)$ is linear. The proof of the Inverted-U property is are simplified by using the signum function

$$\text{sgn}\{w\} = +1 \text{ if } w > 0,\ 0 \text{ if } w = 0, \text{ and } -1 \text{ if } w < 0. \tag{A31}$$

First consider the ON-reaction in equation (A20), which is denoted by x_5 in Figure 8a. Writing the derivative of a function $g(I)$ as $g'(I)$, then, by (A20), for each fixed J,

$$\begin{aligned}\text{sgn}\{x_5'(I)\} = \text{sgn}\{&A^2[f'(I+J) - f'(I)] + 2A[f(I)f'(I+J) - f(I+J)f'(I)] \\ &+ [f^2(I)f'(I+J) - f^2(I+J)f'(I)]\}.\end{aligned} \tag{A32}$$

Since $f(w)$ is sigmoid,

$$f(0) = f'(0) = 0. \tag{A33}$$

Thus, by (A32) and (A33),

$$\text{sgn}\{x_5'(0)\} = \text{sgn}\{A^2 f'(J)\} > 0. \tag{A34}$$

At large values of I,

$$f(I+J) > f(I), \tag{A35}$$

whereas

$$f'(I+J) < f'(I). \tag{A36}$$

Consequently, each term in brackets on the right-hand side of (A32) is negative. Thus, at large I values,

$$\text{sgn}\{x_5'(I)\} < 0 \tag{A37}$$

The inequalities (A34) and (A37) show that, for fixed J, $x_5(I)$ increases and then decreases as a function of I. This is the Inverted-U for the ON-reaction. In fact, since $f(\infty) < \infty$, (A20) implies that $\lim_{I\to\infty} x_5(I) = 0$. A similar proof holds for the OFF-reaction.

A7. Underaroused Hypersensitivity to Suprathreshold Phasic Input Increments. An underaroused syndrome is hypersensitive to phasic input increments that exceed an elevated threshold level. This may be one reason why individuals with autism may react badly to certain otherwise normal sensory stimuli. To prove this property, suppose that I is chosen abnormally small and, consequently, that $f(I)$ is very small because of f's S-shaped graph. Let J represent the intensity of a fearful cue (e.g., a shock level) and let the dipole ON-output (A20) be correlated with the amount of fear. Since I is so small, the "fear threshold is raised" in the sense that a larger value of J is needed to create a large net ON-output than when I is chosen in the "normal" range. Until this output threshold is exceeded, the dipole will exhibit abnormally low responses to fearful cues. On the other hand, although the fear threshold is high, once J is chosen sufficiently large to elicit a detectable net ON-reaction, additional increments in J create larger than normal increments in fear. This is because the terms $f(I)$ in the numerator and denominator of (A20) are abnormally small. More precisely, differentiating (A20) with respect to J, we find the rate at which the ON-output increases to unit increases in J. This rate is

$$\frac{\partial}{\partial J}(S_1 z_1 - S_2 z_2) = \frac{A^2 B f'(I+J)}{[A + f(I+J)]^2}. \tag{A38}$$

If $I + J$ is chosen so that $f(I + J)$ is small but growing rapidly, then $f'(I + J)$ is large when the denominator, $[A + f(I + J)]^2$, is small. In summary, underaroused depression is hyperexcitable above its high threshold.

A8. Paradoxical ON-Response to Small Mismatches and Enhanced Rebounds to Large Mismatches. Two other properties of underaroused dipoles are predicted to be related to the "need for sameness" of an individual with autism; see Section 6.6. These properties, like underaroused hyperexcitability, are due to the faster-than-linear, or threshold, behavior of the S-shaped signal function, $f(w)$, at small activity values, w. Neither property holds if the signal function is linear, say $f(w) = w$. In particular, by (30), when $f(w) = w$, a mismatch-mediated arousal increment ΔI in response to an unexpected event causes a rebound whenever $\Delta I > A$. The minimal ΔI capable of causing a rebound is independent of the ambient arousal level, I. This property does not hold when $f(w)$ grows faster than linearly, say $f(w) = w^2$, which approximates the sigmoid shape of $f(w)$ at low arousal levels. By (A28), a rebound occurs when $f(w) = w^2$ only if

$$\Delta I > g(I, J), \tag{A39}$$

where the function

$$g(I, J) = \frac{A - I(I + J) + (A + I^2)^{1/2}[A + (I + J)^2]^{1/2}}{2I + J} \tag{A40}$$

is a decreasing function of I. In fact, $g(I, J)$ approaches 0 as I is chosen arbitrarily large. Thus, a much larger ΔI is needed to rebound an underaroused dipole than a normally aroused dipole. Moreover, if $\Delta I < AJ^{-1}$, then when $I \cong 0$,

$$\frac{\partial}{\partial(\Delta I)}\left[\frac{(I + \Delta I + J)^2}{A + (I + J)^2} - \frac{(I + \Delta I)^2}{A + I^2}\right] > 0. \tag{A41}$$

In other words, a mismatch-mediated arousal burst can actually enhance the ON-output of an underaroused dipole instead of rebounding the dipole. However, once the rebound threshold is exceeded, then the size of the rebound is greater due to the small arousal terms (i.e., the Weber law) in the denominator of (A28).

A9. Paradoxical Lack of Rebound to Phasic Input Decrement: Ordering of Reinforcement Magnitude. This section predicts how several behavioral indices should all covary as arousal level is parametrically increased. These predictions are important when one considers the hypothesis that certain individuals with autism are underaroused whereas various individuals with schizophrenia are overaroused (Grossberg, 2000b). The first index says that reducing J units of shock (or other negative reinforcer) to $J/2$ units is less rewarding (i.e., produces a smaller rebound) than reducing $J/2$ units of shock to 0 units, despite the fact that both operations reduce shock by $J/2$ units. This result is based on the fact that (A20) and (A25) include Weber law ratios of I and J terms as well as differences of I and J terms. More generally, one can predict when reducing J_1 units of shock to K_1 units at arousal level I_1 is more reinforcing than reducing J_2 units of shock to K_2 units at arousal level I_2 (Grossberg, 1972b). To make these assertions, assume that the size of the relief rebound caused by reducing the shock level is proportional to the rewarding effect of the manipulation, other things being equal.

To simplify the computations, it is convenient to use a signal function

$$f(w) = \max(w - C, 0). \tag{A42}$$

Such a signal function has a threshold C, below which it equals 0 and above which it grows linearly. This threshold function approximates a sigmoid function in the activity range before saturation occurs. Denote the steady-state ON-reaction that occurs after a specific input of intensity J is kept on for S time units by $x_5(S, J \to K)$ and the OFF-rebound that occurs when intensity J is switched to K at time $t = S$ by $x_6(S^+, J \to K)$. To compute $x_6(S^+, J \to K)$, the transmitters z are approximated by their steady-state values at $t = S$ and the potentials x by their new steady-state values in response to input K.

Given an arousal level I that exceeds the threshold, C, then

$$x_6\left(S^+, J \to \frac{J}{2}\right) = \frac{AB\frac{J}{2}(I - A - C)}{(D + I)(D + I + J)}, \tag{A43}$$

where $D = A - C$. By comparison, (A20) and (A25) imply that

$$x_5(S, J \to 0) = \frac{A^2 BJ}{(D + I)(D + I + J)} \tag{A44}$$

and

$$x_6(S^+, J \to 0) = \frac{ABJ(I - C)}{(D + I)(D + I + J)} \tag{A45}$$

from which it also follows that

$$x_6\left(S^+, \frac{J}{2} \to 0\right) = \frac{AB\frac{J}{2}(I - C)}{(D + I)\left(D + I + \frac{J}{2}\right)} \tag{A46}$$

and

$$\frac{x_6(S^+, K \to 0)}{x_5(S, K \to 0)} = A^{-1}(I - C) \tag{A47}$$

for any $K > 0$. Comparing (A43) and (A46), shows that the relative rebound sizes satisfy

$$x_6\left(S^+, \frac{J}{2} \to 0\right) > x_6\left(S^+, J \to \frac{J}{2}\right), \tag{A48}$$

or that cutting J units in half is less rewarding than shutting off $J/2$ units. In addition, the ratio (A47) increases with I, as in the more general equation (A26). Substituting (A47) into (A43) shows that

$$x_6\left(S^+, J \to \frac{J}{2}\right) = \frac{A^2 B\frac{J}{2}[x_5^{-1}(S, K \to 0)x_6(S^+, K \to 0) - 1]}{(D + I)(D + I + J)} \tag{A49}$$

By (A49), an arousal level that favors the possibility of learned avoidance in the presence of fearful cues (i.e., the OFF-rebound is much bigger than the ON-response so that the right hand side of (A49) is positive) also favors a large rewarding effect when the shock level is halved. If, however, I is chosen to be small (underarousal), then x_6 in (A43) can be negative (no rebound occurs) even if x_6 in (A46) is positive (a rebound occurs). In other words, cutting a shock in half may generate no relief, even if shutting off half the shock level totally does cause relief.

APPENDIX B:
SPECTRAL TIMING IN THE START MODEL

START: A Unified Model of Adaptive Timing and Conditioned Reinforcer Learning The START model combines Spectral Timing mechanisms with mechanisms from Adaptive Resonance Theory, or ART (see Section 7), hence its name. The adaptive timing equations described herein model adaptively timed learning in the dentate-CA3 hippocampal circuits; see Figure 12. The later cerebellar adaptive timing model of Fiala, Grossberg, and Bullock (1996) refined adaptive timing concepts to predict how the metabotropic glutamate receptor system may contribute to adaptive timing, notably how mGluR mechanisms may give rise to spectral timing. The present summary provides a simpler analysis of the main adaptive timing concepts and equations.

The START model was tested by simulating parametric data from reinforcement learning experiments, notably classical conditioning experiments. Each sensory event is therefore called a conditioned stimulus, or CS. The i^{th} sensory event is denoted by CS_i. Event CS_i activates a population of cells that is called the i^{th} sensory representation S_i (Figure 9). Another population of cells, called a drive representation D, receives a combination of sensory, reinforcement, and homeostatic (or drive) stimuli. Reinforcement learning, emotional reactions, and motivational decisions are controlled by D. During conditioning, presentation of a CS_i before a US causes activation of S_i followed by activation of D. Such pairing causes strengthening of the adaptive weight, or long term memory trace, in the modifiable synapses from S_i to D. This learning event converts CS_i into a conditioned reinforcer. Conditioned reinforcers hereby acquire the power to activate D via the conditioning process.

In the START model, reinforcement learning in $S_i \rightarrow D$ pathways is supplemented by a parallel learning process T that is concerned with adaptive timing. As shown in Figure 9, both of these learning processes output to D, which in turn inhibits the population of cells A that form the orienting system; also see Figure 2. The orienting system is a source of nonspecific arousal signals that are capable of resetting short-term memory, triggering opponent emotional reactions, attention shifts, and orienting responses. The inhibitory pathway from D to A is a gate that prevents these events from occurring in response to expected disconfirmations (Section 4).

B1. Limited Capacity Short Term Memory The sensory representations S_i compete for a limited capacity, or finite total amount, of activation. Winning populations are stored in short-term memory, or STM. The competition is carried out by an on-center off-surround interaction among the populations S_i. The property of STM storage is achieved by using recurrent, or feedback, pathways among the populations. A tendency to select winning populations is achieved by using membrane equations, or shunting interactions, to define each population's activation, and a proper choice of feedback signals between populations (Grossberg, 1973, 1980). Expressed mathematically, each CS_i activates an STM representation S_i whose activity S_i obeys the shunting on-center off-surround competitive feedback equation:

$$\frac{d}{dt}S_i = -\alpha_A S_i + \beta_A (1 - S_i)(I_i(t) + f_S(S_i)) - \gamma_A S_i \sum_{k \neq i} f_S(S_k). \qquad \text{(B1)}$$

In (B1), $I_i(t)$ is the input that is turned on by presentation of CS_i. Term $-\alpha_A S_i$ describes passive decay of activity S_i. Term $\beta_A(1-S_i)(I_i(t)+f_S(S_i))$ describes the excitatory effect on S_i of the input $I_i(t)$ and the feedback signal $f_S(S_i)$ from population S_i to itself. Activity S_i can continue to grow until it reaches the excitatory saturation point, which is scaled to equal 1 in (1). Term $-\delta_A S_i \sum_{k\neq i} f_S(S_k)$ describes inhibition of S_i by competitive signals $f_S(S_k)$ from the off-surround of populations $k \neq i$. Due to these feedback signals, a brief CS_1 input gives rise to a sustained STM activation S_1, which is partially inhibited by competition from S_0's activation in response to a US. The signal function f_S in (B1) is chosen to suppress noise while contrast enhancing the most active cell activities. In the Grossberg and Merrill (1992) simulations, the simple threshold, or half-wave rectification, function

$$f(w) = [w-\mu]^+ \equiv \max(w-\mu, 0) \tag{B2}$$

was used, except in equation (B8) below, which uses a sigmoid signal function.

B2. Drive Representation The computer simulations reported herein use only a single drive representation D. Explanations of data arising from competing drive representations are discussed in a number of modeling articles; e.g., Grossberg (1972a, 1972b, 1982a, 1984a). The activity D of the drive representation D obeys the equation

$$\frac{d}{dt}D = -\alpha_D D + \beta_D \sum_i f_D(S_i)C_i + \gamma_D R. \tag{B3}$$

In (B3), term $-\alpha_D D$ describes the passive decay of activity D. Term $\beta_D \sum_i f_D(S_i)C_i$ describes the total excitatory effect of all the sensory representations S_i on D. In this term, the signal function f_D is chosen as in (B2), and C_i is the adaptive weight, or long-term memory (LTM) trace, in the pathway from the sensory representation S_i of CS_i to the drive representation D. This LTM trace is denoted by C_i because its size measures how well S_i can activate D, and thus how CS_i $(i \geq 1)$ has become a conditioned reinforcer through learning. Because C_i multiplies $f_D(S_i)$, a large activation of S_i will have a negligible effect on D if C_i is small, and a large effect on D if C_i is large. Coefficient C_0 is set equal to a large value from the start because it enables the US to activate D via its sensory representation S_0. Term $\gamma_D R$ describes the total output of the spectral timing circuit to D. Output R is defined in equation (B11) below.

B3. Conditioned Reinforcement The adaptive weight C_i that calibrates conditioned reinforcement obeys a gated learning law (Grossberg, 1968, 1980):

$$\frac{d}{dt}C_i = \alpha_C S_i(-C_i + \beta_C(1-C_i)f_C(D)). \tag{B4}$$

Learning by C_i is turned on and off by the signal S_i, which thus acts like a learning gate, or modulator. Once turned on, C_i performs a time-average of activity at the drive representation D via the signal $f_C(D)$, which is chosen as in equation (B2). Activity C_1 cannot exceed the finite value 1, due to the shunting term $1-C_i$. The value of C_i can both increase and decrease during

the course of learning. The remaining equations of the model describe the adaptive timing process.

B4. Activation Spectrum The START model is said to control "spectral" timing because, each drive representation D interacts with a population of cell sites whose members react at a spectrum of rates r_j. Neural populations whose elements are distributed along a temporal or spatial parameter are familiar throughout the nervous system. Two examples are populations of spinal cord cells that obey the size principle (Henneman, 1957, 1985), and cells of the visual cortex that are tuned to spatial frequency (Jones and Keck, 1978; Musselwhite and Jeffreys, 1985; Parker and Salzen, 1977a, 1977b; Parker, Salzen, and Lishman, 1982a, 1982b; Plant, Zimmern, and Durden, 1983; Skrandies, 1984; Vassilev, Manahilov, and Mitov, 1983; Vassilev and Strashimirov, 1979; Williamson, Kaufman, and Brenner, 1978). The spectral activities x_{ij} that are associated with drive representation D and activated by sensory representation S_i obey the equation

$$\frac{d}{dt}x_{ij} = r_j(-x_{ij} + (1-x_{ij})f_x(S_i)), \tag{B5}$$

where f_x satisfies equation (B2). By equations (B1) and (B5), presentation of CS_i to S_i via an input I_i generates an output signal $f_x(S_i)$ that activates the local potentials x_{ij} of all cell sites in the target population. The potentials x_{ij} respond at rates proportional to $r_j, j = 1, 2, ..., n$, where j indexes the different cell sites that form the spectrum. These potentials activate the next processing stage via signals

$$f(x_{ij}) = \frac{x_{ij}^8}{\delta_{ij}^8 + x_{ij}^8}. \tag{B6}$$

Signal $f(x_{ij})$ is a sigmoid function of activity x_{ij}. Figure 12a shows the activation spectrum $f_{ij} = f(x_{ij}(t))$ that arises from presentation of CS_i to S_i via input I_i in equation (B1), using a choice of rate parameters r_j in equation (B5) which range from 10 (fast) to 0.0025 (slow).

B5. Habituative Transmitter Spectrum Each spectral activation signal $f(x_{ij})$ interacts with a habituative chemical transmitter y_{ij} via the equation

$$\frac{d}{dt}y_{ij} = \alpha_y(1-y_{ij}) - \beta_y f(x_{ij})y_{ij}. \tag{B7}$$

This is the same equation for transmitter habituation as in equations (A14) and (A15) of Appendix A. According to equation (B7), the amount of neurotransmitter y_{ij} accumulates to a constant target level 1, via term $\alpha_y(1-y_{ij})$, and is inactivated, or *habituates*, due to a mass action interaction with signal $f(x_{ij})$, via term $-\beta_y f(x_{ij})y_{ij}$. The different rates r_j at which each x_{ij} is activated causes the corresponding y_{ij} to become habituated at different rates as well. The family of curves $y_{ij}(t)$, $j = 1, 2, ..., n$, is called a habituation spectrum. The signal functions $f(x_{ij}(t))$ in Figure 12a generate the habituation spectrum of $y_{ij}(t)$ curves in Figure 12b.

B6. Gated Signal Spectrum Each signal $f(x_{ij})$ interacts with y_{ij} via mass action to generate a net output signal from its population of cell sites. This process is also called *gating* of $f(x_{ij})$ by

y_{ij} to yield a net output signal g_{ij} that is equal to $f(x_{ij})y_{ij}$. Each gated signal $g_{ij}(t) \equiv f(x_{ij}(t))y_{ij}(t)$ has a different rate of growth and decay, thereby generating the gated signal spectrum shown in Figure 12c. Each of the functions $g_{ij}(t)$ is a unimodal function of time that is maximally positive at different times. Each function $g_{ij}(t)$ achieves its maximum value M_{ij} at time T_{ij}, where T_{ij} is an increasing function of j, and M_{ij} is a decreasing function of j.

B7. Spectral Learning Law Learning of spectral timing obeys an equation

$$\frac{d}{dt} z_{ij} = \alpha_z f(x_{ij}) y_{ij} (-z_{ij} + N), \tag{B8}$$

where N is a transient Now Print signal (see Section B8) that is derived from the activity D of the drive representation in equation (B3). Activity of D is, in turn, caused by reinforcing events such as a US. Each *long-term memory (LTM) trace* z_{ij} in equation (B8) is computed at the end of the pathway, or synapse, that processes the gated signal $g_{ij} = f(x_{ij})y_{ij}$. The signal g_{ij} acts as a sampling signal that turns on, or *gates*, the learning process, and causes z_{ij} to approach N during the sampling interval at a rate proportional to g_{ij}. The attraction of z_{ij} to N is called *steepest descent*. Thus (8) is an example of learning by *gated steepest descent*. Each z_{ij} changes by an amount that reflects the degree to which the curves $g_{ij}(t)$ and $N(t)$ have simultaneously large values through time. If g_{ij} is large when N is large, then z_{ij} increases in size. If g_{ij} is large when N is small, then z_{ij} decreases in size. As in equation (B4), z_{ij} can either increase or decrease as a result of learning. Because the different g_{ij} peak at different times, the entire population can work together to learn the interstimulus intervals, or ISIs, between the CS and US.

Associative learning by gated steepest descent was introduced into neural network models in Grossberg (1969a) and is the learning law that was used to introduce Adaptive Resonance Theory (Grossberg, 1976a, 1976b). An associative learning law of this form was subsequently used by Levy and his colleagues to model their data on hippocampal Long Term Potentiation (LTP) and Long Term Depression (LTD) (Levy, Brassel, and Moore, 1983, Levy and Desmond, 1985). Singer (1983) has also used such a law to model his experiments on adaptive tuning of visual cortical cells during the visual critical period. These experiments support the Adaptive Resonance Theory predictions (Grossberg, 1976a, 1976b, 1980) that both hippocampal LTP/LTD and feature detector tuning in visual cortex should obey a learning law with gating properties that is capable of both increasing (LTP) and decreasing (LTD) synaptic weights during associative learning.

B8. Now Print Signal A transiently active Now Print signal N modulates the learning process in equation (B8). The signal N may be activated either by a US or by a CS that has already become a conditioned reinforcer. Both the US and a conditioned reinforcer CS can activate the drive representation D, as shown in equation (B3). We assume that the Now Print signal N is turned on by sufficiently large and rapid increments in the activity D. Such a transient signal N may be derived from a change in sustained activity D by the action of a slow inhibitory interneuron, as in equation

$$N = [f_C(D) - E - \varepsilon]^+. \tag{B9}$$

In equation (B9), the signal $f_C(D)$ from the drive representation D attempts to excite the site where learning occurs. It is balanced by E, which is the activity of an inhibitory interneuron that more slowly time-averages $f_C(D)$, as in equation

$$\frac{d}{dt}E = \alpha_E(-E + f_C(D)). \tag{B10}$$

The notation $[w-\varepsilon]^+$ in equation (B9) means that ε is an output threshold for the signal $w = f_C(D) - E$. In other words, $N = 0$ if $f_C(D) - E \le \varepsilon$, and $N = f_C(D) - E - \varepsilon$ if $f_C(D) - E > \varepsilon$. The Now Print signal N responds to increments in D with brief learning signals that occur when the US turns on and that can be sampled by the spectrally timed signals g_{ij}. An important property of N is that it increases in amplitude, but not significantly in duration, in response to larger inputs $f_C(D)$. Thus learning is faster in response to stronger rewards.

B9. Doubly Gated Signal Spectrum Each long-term memory trace z_{ij} learns to a different degree depending upon how well its spectral rate coincides with the interstimulus interval between CS onset and US onset. These learned traces, in turn, gate the signals $g_{ij}(t)$ in order to generate a twice-gated output signal $h_{ij}(t) = f(x_{ij}(t))y_{ij}(t)z_{ij}(t)$ from each of the differently timed cell sites. Each twice-gated signal function $h_{ij}(t)$ thus registers how well the timing of CS and US is learned and read-out by the j^{th} spectral rate. Figure 12d plots the output signals $h_{ij}(t)$. Comparison with the corresponding signals $g_{ij}(t)$ in Figure 12c shows how adaptively timed learning changes the relative strength of each spectral output without changing its timing.

B10. Output Signal The total output R of the network is the sum of the twice-gated signals $h_{ij}(t)$ from all the spectral components corresponding to all the CS_i. Thus

$$R = \sum_{i,j} f(x_{ij})y_{ij}z_{ij}. \tag{B11}$$

The output signal R computes the cumulative learned reaction of the whole population to the input pattern. Figure 12e shows the function R derived from the h_{ij} shown in Figure 12d. A comparison of Figures 12c-e illustrates how the output $R(t)$ generates an accurately timed response from the cumulative partial learning of all the cell sites in the population spectrum. The once-gated signals $g_{ij}(t)$ in Figure 12c are biased towards early times. The twice-gated signals $h_{ij}(t)$ in Figure 12d are biased towards the ISI, but many signals peak at other times. The output $R(t)$ combines these partial views into a cumulative response that peaks at the ISI. Adaptively timed learning is thus a property of this entire population of cell sites.

REFERENCES

Abbott, L.F., Varela, K., Sen, K. and Nelson, S.B. (1997). Synaptic depression and cortical gain control. *Science*, 275:220–223.

Adolphs, R., Tranel, D., and Damasio, A.R. (1998). The human amygdala in social judgement. *Nature*, 393:470–474.

Aggleton, J.P. (1993). The contribution of the amygdala to normal and abnormal emotional states. *Trends in Neurosciences*, 16:328–333.

Albus, J.S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, 10:25–61.

Allen, G., Buxton, R.B., Wong, E.C., and Courchesne, E. (1997). Attentional activation of the cerebellum independent of motor involvement. *Science*, 275:1940–1943.

Allen, G., and Courchesne, E. (2001). Attention function and dysfunction in autism. *Frontiers in Bioscience*, 6:105–119.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, 4th edition.* Washington, DC: American Psychiatric Press.

Ahissar, M. and Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Science*s *USA*, 90:5718–5722.

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

Aylward, E.H., Minshew, N.J., Goldstein, G., Honeycutt, N.A., Augustine, A.M., Yates, K.O., Barta, P.E., and Pearlson, G.D. (1999). MRI volumes of amygdala and hippocampus in non-mentally retarded autistic adolescents and adults. *Neurology*, 53:2145–2150.

Bachevalier, J., and Loveland, K. (2003). Early medial temporal dysfunction and autism. In D. Cicchetti, and E.F. Walker, (Eds.); *Neurodevelopmental Mechanisms in Psychopathology*. Cambridge:Cambridge University Press.

Bailey, A., Luthert, P., Dean, A., Harding, B., Janota, A., Montgomery, M., Rutter, M., and Lantos, P. (1998). A clinicopathophysiologic study of autism. *Brain*, 121:889–905.

Baloch, A. and Waxman, A. (1991). Visual learning, adaptive expectations, and behavioral conditioning of the mobile robot MAVIN. *Neural Networks*, 4:271–302.

Banquet, J.-P., and Grossberg, S. (1987). Probing cognitive processes through the structure of event-related potentials during learning: An experimental and theoretical analysis. *Applied Optics*, 26:4931–4946.

Baranek, G.T. (1999). Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors. *Journal of Autism and Developmental Disorders*, 29:213–224.

Barbas, H. (1995). Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, 19:499–510.

Baron-Cohen, S. (1989). The autistic child's theory of the mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, and Allied Disciplines,* 30:285–297.

Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T., Brammer, M.J., and Simmons, A. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *The European Journal of Neuroscience*, 11(6):1891–1898.

Bauman, M.L., and Kemper, T.L. (1986). Developmental cerebellar abnormalities: A consistent finding in early infantile autism. *Neurology*, 36:190.

Baxter, M.G., Parker, A., Lindner, C.C.C., Izquierdo, A.D., and Murray, E.A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *Journal of Neuroscience*, 20:4311–4319.

Belmonte, M.K., Cook, E.H. Jr., Anderson, G.M., Rubenstein, J.L., Greenough, W.T., Beckel-Mitchener, A., Courchesne, E., Boulanger, L.M. Powell, S.B., Levitt, P.R., Perry, E.K., Jiang, Y.H., DeLorey, T.M., and Tierney, E. (2004). Autism as a disorder of neural information processing: directions for research and targets for therapy (1). *Molecular Psychiatry,* 9(7):646–663.

Berger, T.W., Berry, S.D., and Thompson, R.F. (1986). Role of the hippocampus in classical conditioning of aversive and appetitive behaviors. In R.L. Isaacson and K.H. Pribram, (Eds.), *The Hippocampus, Volume 4.* New York:Plenum Press, pp. 203–239.

Beversdorf, D.Q., Anderson, J.M., Manning, S.E., Anderson, S.L., Nordgren, R.E., Felopulos, G.J., and Bauman, M.L. (2001). Brief Report: Macrographia in high functioning adults with autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, 31:97–101.

Blumer, D. and Benson, D.F. (1975). Personality changes with frontal lobe lesions. In D.F. Benson and D. Blumer (Eds.) *Psychiatric Aspects of Neurological Disease*. New York: Grune and Stratton, pp 151-170.

Bower, G.H. (1981). Mood and memory. *American Psychologist* 36:129–148.

Bowler, D.M. (1992). "Theory of the mind" in Asperger's syndrome. *Journal of Child Psychology and Psychiatry,* 33:877–893.

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

Brown, J., Bullock, D., and Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience,* 19:10502–10511.

Brown, J., Bullock, D., and Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, 2004, 17:471–510.

Bruner, J.S. (1969). Modalities of memory. In G.A. Talland and N.C. Waugh, (Eds.), *The Pathology of Memory*. New York:Academic Press, pp. 253–259.

Bullier, J., Hupé, J.M., James, A. and Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology (Paris)*, 90:217–220.

Bullock, D., Cisek, P., and Grossberg, S. (1998). Cortical networks for control of voluntary arm movements under variable force conditions. *Cerebral Cortex* 8:48–62.

Bullock, D., Fiala, J.C., and Grossberg, S. (1994). A neural model of timed response learning in the cerebellum. *Neural Networks*, 7:1101–1114.

Caputo, G. and Guerra, S. (1998). Attentional selection by distractor suppression. *Vision Researc*h, 38:669–689.

Carpenter, G.A. (2001). Neural-network models of learning and memory: leading questions and an emerging framework. *Trends in Cognitive Sciences*, 5:114–118.

Carpenter, G.A. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37:54–115.

Carpenter, G.A. and Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, 3:129–152.

Carpenter, G.A., and Grossberg, S. (1991). *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA:MIT Press.

Carpenter, G.A., and Grossberg, S. (1993). Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences,* 16:131–137.

Casanova, M.F., Buxhoeveden, D.P., Switela, A.E., and Roy, E. (2002). Minicolumnar pathology in autism. *Neurology*, 58(3):428–432.

Chang, C. and Gaudiano, P. (1998). Application of biological learning theories to mobile robot avoidance and approach behaviors. *Journal of Complex Systems*, 1:79–114.

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., and Drew, A. (1997). Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology*, 33:781–789.

Chugani, D.C., Sundram, B.S., Beham, M., Lee, M., and Moore, G.J. (1999). Evidence of altered energy metabolism in autistic children. *Progress in Neuropsychopharmocology and Biological Psychiatry,* 23:635–641.

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

Cohen, N.J., and Squire, L.R. (1980). Preserved learning and retention of a pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210:207–210.

Contreras-Vidal, J.L., Grossberg, S., and Bullock, D. (1997). A neural model of cerebellar learning for arm movement control: Cortico-spinal-cerebellar dynamics. *Learning and Memory* 3:475–502.

Courchesne, E., Carper, R., and Akshoomoff, N. (2003). Evidence of brain overgrowth in the first year of life in autism. *Journal of the American Medical Association*, 290:337–344.

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

Courchesne, E., Townsend, and J., Saitoh, O. (1994). The brain in infantile autism: Posterior fossa structures are abnormal. *Neurology*, 44:214–223.

Courchesne, E., Yeung-Courchesne, R., Press, G.A., Hesselink, J.R., and Jernigan, T.L. (1988). Hypoplasia of cerebellar lobules VI and VII in autism. *New England Journal of Medicine*, 318:1349–1354.

Cox, A., Klein, K., and Charman, T. (1999). The early diagnosis of autism spectrum disorders: Use of the autism diagnostic interview-revised at 20 months and 42 months of age. *Journal of Child Psychology and Psychiatry, and Allied Disciplines,* 40:705–718.

Critchley, H.D., Daly, E.M., Bullmore, E.T., Williams, S.C.R., Van Amelsvoort, T., Robertson, D.M., Rowe, A., Phillips, M., McAlonan, G., Howlin, P., and Murphy, D.G.M. (2000). The functional neuroanatomy of social behaviour: Changes in cerebral blood flow when people with autistic disorder process facial expressions. *Brain,* 123:2203–2212.

Davis, M. (1994). The role of the amygdala in emotional learning. *International Review of Neurobiology* 36:225–265.

Dawson, G., Finley, C., Phillips, S., Galpert, L., and Lewy, A. (1988). Reduced P3 amplitude of the event-related brain potential: Its relationship to language ability in autism. *Journal of Autism and Developmental Disorders*, 18:493–504.

Damasio, A.R., Tranel, D., Damasio, H. (1991). Somatic markers and the guidance of behavior: theory and preliminary testing. In H.S. Levin, H.M. Eisenberg, A.L. Benton, (Ed.s). *Frontal Lobe Function and Dysfunction*. Oxford: Oxford University Press, pp 217-229.

Dawson, G., Munson, J., Estes, A., Osterling, J., McPartland, J., Toth, K., Carver, L., and Abbott, R. (2002). Neurocognitive function and joint attention ability in young children with autistic spectrum disorder versus developmental delay. *Child Development*, 73:345–358.

Deadwyler, S.A., West, M.O., and Lynch, G. (1979). Activity of dentate granule cells during learning: Differentiation of perforant path inputs. *Brain Research*, 169:29–43.

Deadwyler, S.A., West, M.O., and Robinson, J.H. (1981). Entorhinal and septal inputs differentially control sensory-evoked responses in the rat dentate gyrus. Science, 1181–1183.

De Fosse, L., Hodge, S.M., Makris, N., Kennedy, D.N., Caviness, V.S. Jr., McGrath, L., Steele, S., Ziegler, D.A., Herbert, M.R., Frazier, J.A., Tager-Flusberg, H., and Harris, G.J. (2004). Language-association cortex asymmetry in autism and specific language impairment. *Annals of Neurology,* 56(6):755–756.

DeLong, G.R., and Heinz, E.R. (1997). The clinical syndrome of early-life bilateral hippocampal sclerosis. *Annals of Neurology*, 42:11–17.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, 3:1–8.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London,* 353:1245–1255.

Desimone, R., and Ungerleider, L.G. (1989). Neural mechanisms of visual processing in monkeys. In: F. Boller and J. Grafman, (Eds.), *Handbook of Neuropsychology*, *Volume 2.* Amsterdam:Elsevier, pp. 267–299.

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (1994). American Psychiatric Association.

Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psycholog*y: *Human Perception and Performanc*e, 14:188–202.

Ebner, T.J., and Bloedel, J.R. (1981). Correlation between activity of Purkinje cells and its modification by natural peripheral stimuli. *Journal of Neurophysiology*, 45:948–961.

Eichenbaum, H., Otto, T., and Cohen, N.J. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, 17:449–472.

Eigsti, I.M., and Shapiro, T. (2003). A systems neuroscience approach to autism: biologic, cognitive, and clinical perspectives. *Mental Retardation and Developmental Disabilities Research Reviews,* 9(3):205–215.

Engel, A. K., Fries, P. and Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience,* 2:704–716.

Ersoy, B., Carpenter, G.A., and Grossberg, S. (2002). Top-down expectations during cortical learning of recognition categories. *Society for Neuroscience Abstracts,* 872.1.

Evarts, E.V. (1973). Motor cortex reflexes associated with learned movement. *Science* 179:501–503.

Fatemi, S.H., Halt, A.R., Realmuto, G., Earle, J., Kist, D.A., Thuras, P., and Merz, A. (2002). Purkinje cell size is reduced in cerebellum of patients with autism. *Cellular and Molecular Neurobiology,* 22:171–175.

Felleman, D.J., and van Essen, C.D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1-47.

Fiala, J.C., Grossberg, S., and Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye blink response. *Journal of Neuroscience*, 16:3760–3774.

Filipek, P.A., Accardo, P.J., Ashwal, S., Baranek, G.T., Cook, E.H. Jr, Dawson, G., Gordon, B., Gravel, J.S., Johnson, C.P., Kallen, R.J., Levy, S.E., Minshew, N.J., Ozonoff, S., Prizant, B.M., Rapin, I., Rogers, S.J., Stone, W.L., Teplin, S.W., Tuchman, R.F., and Volkmar, F.R. (2000). Practice parameter: Screening and diagnosis of autism: Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology*, 55:468–479.

Finch, E.A., and Augustine, G.J. (1998). Local calcium signalling by inositol-1,4,5-triphosphate in Purkinje cell dendrites. *Nature*, 396:753–756.

Fombonne, E. (1999). The epidemiology of autism: a review. *Psychological Medicine,* 29:769–786.

Fox, K., Schlaggar, B.L., Glazewski, S., and O'Leary, D.D. (1996). Glutamate receptor blockade at cortical synapses disrupts development of thalamocortical and columnar organization in somatosensory cortex. *Proceedings of the National Academy of Sciences USA*, 93:5584–5589.

French, R.M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3:128–135.

Fries, P., Reynolds, J.H., Rorie, A.E and Desimone, R. (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science,* 291:1560–1563.

Frith, U., and Frith, C.D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358:459–473.

Frith, U., and Hill, E. (2003). *Autism: Mind and Brain*. Oxford, UK:Oxford University Press.

Fujita, M. (1982a). Adaptive filter model of the cerebellum. *Biological Cybernetics*, 45:195–206.

Fujita, M. (1982b). Simulation of adaptive modification of the vestibulo-ocular reflex with an adaptive filter model of the cerebellum. *Biological Cybernetics*, 45:207–214.

Fulton, J.F. (1950). *Frontal Lobotomy and Affective Behavior.* New York: Norton.

Fuster, J.M. (1989). *The Prefrontal Cortex* (second edition). New York: Raven Press.

Gaffan, D. (1974). Recognition impaired and association intact in the memory of monkeys after transection of the fornix. *Journal of Comparative and Physiological Psychology*, 86:1100–1109.

Gaffan, D. (1992). Amnesia for complex naturalistic scenes and for objects following fornix transection in the rhesus monkey. *European Journal of Neuroscience*, 4:381–388.

Gao, E. and Suga, N. (1998). Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proceedings of the National Academy of Sciences* 95:12663–12670.

Gaudiano, P. and Chang, C. (1997). Adaptive obstacle avoidance with a neural network for operant conditioning: Experiments with real robots. *Proceedings of the 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 13–18.

Gaudiano, P., Zalama, E., Chang, C. and Lopez-Coronado, J. (1996). A model of operant conditioning for adaptive obstacle avoidance. In P. Maes *et al.* (Eds.). *From Animals to*

Animats 4. Proceedings of the fourth International Conference on Simulation of Adaptive Behavior, 373–38l. Cambridge, MA: MIT Press.

Ghaziuddin, M., and Butler, E. (1998). Clumsiness in autism and Asperger syndrome: A further report. *Journal of Intellectual Disabilities Research*, 42:43–48.

Gilbert, P.F.C., and Thach, W.T. (1977). Purkinje cell activity during motor learning. *Brain Research*, 128:309–328.

Gloor, P., Olivier, A., Quesney, L.F., Andermann, F., and Horowitz, S. (1982). The role of the limbic system in experiential phenomena of temporal lobe epilepsy. *Annals of Neurology*, 12:129–144.

Gochin, P.M., Miller, E.K., Gross, C.G., and Gerstein, G.L. (1991). Functional interactions among neurons in inferior temporal cortex of the awake macaque. *Experimental Brain Research*, 84:505–516.

Goldberg, M.C., Lasker, A.G., Zee, D.S., Garth, E., Tien, A., and Landa, R.J. (2002). Deficits in the initiation of eye movements in the absence of a visual target in adolescents with high functioning autism. Neuropsychologia, 40: 2039-2049.

Goldstein, S., and Schebach, A.J. (2004). The comorbidity of Pervasive Developmental Disorder and Attebtion Disorder: results of a retrospective chart review. (2004). *Journal of Autism and Developmental Disorders*. 34(3):329–339.

Gove, A., Grossberg, S., and Mingolla, E. (1995). Brightness perception, illusory contours, and corticogeniculate feedback. *Visual Neuroscience,* 12:1027–1052.

Griffen, E.M., Pennington, B.F., Wehner, E.A., and Rogers, S.J. (1999). Executive functions in young children with autism. *Child Development*, 70:817–832.

Grossberg, S. (1968). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences*, 60:758–765.

Grossberg, S. (1969a). On learning and energy-entropy dependence in recurrent and nonrecurrent signed networks. *Journal of Statistical Physics*, 1: 319–350.

Grossberg, S. (1969b). On learning of spatiotemporal patterns by networks with ordered sensory and motor components, I: Excitatory components of the cerebellum. *Studies in Applied Mathematics*, 48:105–132.

Grossberg, S. (1969c). On the production and release of chemical transmitters and related topics in cellular control. *Journal of Theoretical Biology*, 22:325–364.

Grossberg, S. (1971). On the dynamics of operant conditioning. *Journal of Theoretical Biology*, 33:225-255.

Grossberg, S. (1972a). A neural theory of punishment and avoidance, I: Qualitative theory. *Mathematical Biosciences,* 15:39–67.

Grossberg, S. (1972b). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences*, 15:253–285.

Grossberg, S. (1972c). Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes. *Kybernetik*, 10:49–57.

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52:217–257. Reprinted in Grossberg, S. (1982). *Studies of Mind and Brain.* Amsterdam: Kluwer.

Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23:121–134.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23: 187–202.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review,* 87:1–51.
Grossberg, S. (1982a). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review,* 89:529–572.
Grossberg, S. (1982b). *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control.* Hingham, MA:Reidel/Kluwer.
Grossberg, S. (1984a). Some psychophysiological and pharmacological correlates of a developmental, cognitive and motivational theory. *Annals Of The New York Academy Of Sciences,* 425:58–151.
Grossberg, S. (1984b). Some normal and abnormal behavioral syndromes due to transmitter gating of opponent processes. *Biological Psychiatry,* 19:1075–1118.
Grossberg, S. (1987). *The Adaptive Brain, Volume I.* New York:Elsevier.
Grossberg, S. (1988) Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1:17–61.
Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision,* 12:163–185.
Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition,* 8:1–44.
Grossberg, S. (2000a). How hallucinations may arise from brain mechanisms of learning, attention, and volition. *Journal of the International Neuropsychological Society*, 6:583–592.
Grossberg, S. (2000b). The imbalanced brain: From normal behavior to schizophrenia. *Biological Psychiatry,* 48:81–98.
Grossberg, S. (2003a). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews*, 2:47–76.
Grossberg, S. (2003b). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31:423–445.
Grossberg, S., Boardman, I., and Cohen, M.A. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance,* 23:481–503.
Grossberg, S., Carpenter, G.A., and Ersoy, B. (2005). Brain categorization: learning, attention, and consciousness. *Proceedings of the 2005 International Joint Conference on Neural Networks (IJCNN'05)*, in press.
Grossberg, S. and Grunewald, A. (1997). Cortical synchronization and perceptual framing. *Journal of Cognitive Neuroscience*, 9:117–132.
Grossberg, S., and Gutowski, W. (1987). Neural dynamics of decision making under risk: Affective balance and cognitive-emotional interactions. *Psychological Review*, 94:300–318.
Grossberg, S., and Kuperstein, M. (1986). *Neural Dynamics of Adaptive Sensory-Motor Control: Ballistic Eye Movements.* New York:Elsevier.
Grossberg, S., and Levine, D.S. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: Blocking, inter-stimulus interval, and secondary reinforcement. *Applied Optics*, 26:5015–5030.
Grossberg, S., and Merrill, J.W.L. (1992). A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. *Cognitive Brain Research*, 1:3–38.

Grossberg, S., and Merrill, J.W.L. (1996). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. *Journal of Cognitive Neuroscience*, 8:257–277.

Grossberg, S., and Myers, C.W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review,* 107:735–767.

Grossberg, S., and Paine, R.W. (2000). A neural model of cortico-cerebellar interactions during attentive imitation and predictive learning of sequential handwriting movements. *Neural Networks,* 13:999–1046.

Grossberg, S., and Schmajuk, N.A. (1987). Neural dynamics of attentionally-modulated Pavlovian conditioning: Conditioned reinforcement, inhibition, and opponent processing. *Psychobiology,* 15:195–240.

Grossberg, S., and Schmajuk, N.A. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks*, 2:79–102.

Grossberg, S. and Somers, D. (1991). Synchronized oscillations during cooperative feature linking in a cortical model of visual perception. *Neural Networks*, 4:453–466.

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

Gustafsson, L., (2004). Comment on "Disruption in the inhibitory architecture of the cell minicolumn: implications for autism", *Neuroscientist*, 10:189–91.

Hadjikhani, N., Joseph, R.M., Snyder, J., Chabris, C.F., Clark, J., Steele, S., McGrath, L., Vangel, M., Aharon, I., Feczko, E., Harris, G.J., and Tager-Flusberg, H. (2004). Activation of the fusiform gyrus when individuals with autism spectrum disorders view faces. *Neuroimage*, 22:1141–1150.

Halgren. E., Walter, R.D., Cherlow, D.G., and Crandall, P.H. (1978). Mental phenomena evoked by electrical stimulations of the human hippocampal formation and amygdala. *Brain* 101:83–117.

Hallett, M., Lebiedowska, M.K., Thomas, S.L., Stanhope, S.J., Denckla, M.B., and Rumsey, J. (1993). Locomotion of autistic adults. *Archives of Neurology*, 50:1304–1308.

Happe, F. (1996). Studying weak central coherence at low levels: Children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines,* 37970:873–877.

Happe, F. and Frith, U. (1996). The neuropsychology of autism. *Brain*. 119(4):1377-400.

Hardan, A.Y., Minshew, N.J., Harenski, K., and Keshavan, M.S. (2001). Posterior fossa magnetic resonance imaging in autism. *Journal of the American Academy of Child and Adolescent Psychiatry,* 40:666–672.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.

Harries, M.H., and Perrett, D.I. (1991). Visual processing of faces in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Journal of Cognitive Neuroscience*, 3:9–24.

Harris, N.S., Courchesne, E., Townsend, J., Carper, R.A., and Lord, C. (1999). Neuroanatomic contributions to slowed orienting of attention in children with autism. *Cognitive Brain Research,* 8:61–71.

Henneman, E. (1957). Relation between size of neurons and their susceptibility to discharge. *Science*, 26: 1345–1347.

Henneman, E. (1985). The size-principle: A deterministic output emerges from a set of probabilistic connections. *Journal of Experimental Biology*, 115: 105–112.

Herbert, M.R. (2004). Neuroimaging in disorders of social and emotional functioning: what is the question? *Journal of Child Neurology,* 19:772–784.

Herbert, M.R., Ziegler, D.A., Deutsch, C.K., O'Brien, L.M., Lange, N., Bakardjiev, A., Hodgson, J., Adrien, K.T., Steele, S., Makris, N., Kennedy, D., Harris, G.J., and Caviness, V.S. Jr. (2003). Dissociations of cerebral cortex, subcortical and cerbral white matter volumes in autistic boys. *Brain,* 126:1182–1192.

Herbert, M.R., Ziegler, D.A., Deutsch, C.K., O'Brien, L.M., Kennedy, D., Filipek, P.A., Bakardjiev, A.I., Hodgson, J., Takeoka, A., Makris, N., and Caviness, V.S. Jr. (2005). Brain asymmetries in autism and developmental language disorder: a nested whole-brain analysis. *Brain,* 128:213–226.

Herbert, M.R., Ziegler, D.A., Makris, N., Filipek, P.A., Kemper, T.L., Normandin, J.J., Sanders, H.A., Kennedy, D.N., and Caviness, V.S. Jr. (2004). Localization of white matter volume increase in autism and developmental language disorder. *Annals of Neurology,* 55:530–540.

Hoehler, F.K. and Thompson, R.F. (1980). Effect of the interstimulus (CS-UCS) interval on hippocampal unit activity during classical conditioning of the nictitating membrane response of the rabbit (Oryctolagus cuniculus). *Journal of Comparative and Physiological Psychology*, 94:201-215.

Howard, M.A., Cowell, P.E., Boucher, J., Broks, P., Mayes, A., Farrant, A., and Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. *Neuroreport*, 11:2931–2935.

Hubel, D.H., and Wiesel, T.N. (1965). Binocular interaction in the striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology*, 28:1041–1059.

Hubl, D., Bolte, S., Feineis-Matthews, S., Lanfermann, H., Federspeil, A., Strik, W., Poustka, F., and Dierks, T. (2003). Functional imbalance of visual pathways indicates alternative face processing strategies in autism. *Neurology,* 61:1232–1237.

Hughs, C., Russell, J., and Robbins, T. (1996). Evidence for executive dysfunction in autism. *Brain,* 119:1377–1400.

Ichise, T., Kano, M., Hashimoto, K., Yangihara, D., Nakao, K., Shigemoto, R., Katsuki, M., and Aiba, A. (2000). mGluR1 in cerebellar Purkinje cells essential for long-term depression, synapse elimination, and motor coordination. *Science*, 288:1832–1835.

International Classification of Diseases (1994). In Diagnostic Criteria for Research, Tenth Edition. World Health Organization: Geneva.

Ito, M. (1984). *The Cerebellum and Neural Control*. New York:Raven Press.

Jolliffe, T., and Baron-Cohen, S. (1997). Are people with autism and Asperger syndrome faster than normal on the embedded figure test? *Journal of Child Psychology and Psychiatry*, 38:527–534.

Jones, R., and Keck, M.J. (1978). Visual evoked response as a function of grating spatial frequency. *Investigative Ophthalmology and Visual Science*, 17: 652–659.

Jones, V., and Prior, M.J. (1985). Motor imitation abilities and neurological signs in autistic children. *Journal of Autism and Developmental Disorders*, 15:37–46.

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

Kalaska, J.F., Cohen, D.A.D., Hyde, M.L., and Prud'homme, M.J. (1989). A comparison of movement direction-related versus load direction-related activity in primate motor cortex using a two-dimensional reaching task. *Journal of Neuroscience* 9:2080–2102.

Kanner, L. (1943). Autistic disturbance of affective content. *The Nervous Child*, 2:217–250.

Kaas, J. H. (1999). Is most of neural plasticity in the thalamus cortical? *Proceedings of the National Academy of Sciences USA,* 96:7622–7623.

Kastner, S. and Ungerleider, L.G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia,* 39:1263–1276.

Kemper, T.L., and Bauman, M.L. (1993). The contribution of neuropathologic studies to the understanding of autism. *Behavioral Neurology*, 11:175–187.

Knowlton, B.J. and Squire, L.R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262(5140):1747–1749.

Koshino, H., Carpenter P.A., Minshew, N.J., Cherkassky, V.L., Keller, T.A., and Just, M.A. (2005). Functional connectivity in an fMRI working memory task in high-functioning autism. *Neuroimage,* 24:810–821.

Krupa, D.J., Ghazanfar, A.A. and Nicolelis, M.A. (1999). Immediate thalamic sensory plasticity depends on corticothalamic feedback. *Proceedings of the National Academy of Sciences USA,* 96:8200–8205.

LeDoux, J.E. (1993). Emotional memory systems in the brain. *Behavioural Brain Research* 58:69–79.

Lainhart, J.E., Piven, J., Wzorek, M., Landa, R., Santangelo, S.L., Coon, H., and Folstein, S.E. (1997). Macrocephaly in children and adults with autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36:282–290.

Lamb, J.A., Moore, J., Bailey, A., and Monaco, A.P. (2000). Autism: Recent molecular genetic advances. *Human Molecular Genetics*, 9:861–868.

Lee, D.A., Lopez-Alberola, R., and Bhattacharjee, M. (2003). Childhood autism: a circuit syndrome? *Neurologist,* 9:99–109.

Le, T.H., Pardo, J.V., and Hu, X. (1998). 4 T-fMRI study of nonspatial shifting of selective attention: Cerebellar and parietal contributions. *Journal of Neurophysiology*, 79:1535–1548.

Levisohn, L., Cronin-Golomb, A., and Schwahmann, J.D. (2000). Neuropsychologic consequences of tumor resection in children with cerebellar cognitive affective syndrome in a paediatric population. *Brain,* 123:1041–1050.

Levy, W.B., Brassel, S.E., and Moore, S.D. (1983). Partial quantification of the associative synaptic learning rule of the dentate gyrus. *Neuroscience*, 81: 799–808.

Levy, W.B. and Desmond, N.L. (1985). The rules of elemental synaptic plasticity. In: *Synaptic modification, neuron selectivity and nervous system organization* (W.B. Levy, J. Anderson, and S. Lehmkuhle, Eds.). Hillsdale, NJ: Erlbaum Associates, pp.105–121.

Liddle PF (1994): Volition and schizophrenia. In A.S. David and J.C Cutting (Eds.). *The Neuropsychology of Scizophrenia.* Hillsdale: Erlbaum Press, pp 39-49.

Lisberger, S.G. (1988). The neural basis for motor learning in the vestibulo-ocular reflex in monkeys. *Trends in Neurosciences*, 11:147–152.

Luck, S. J., Chelazzi, L., Hillyard, S. A. and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiolog*y, 77:24–42.

Luna, B., Minshew, N.J., Garver, K.E., Lazar, N.A., Thulborn, K.R., Eddy, W.F., and Sweeny, J.A. (2002). Neocortical system abnormalities in autism: An fMRI study of spatial working memory. *Neurology*, 59:834–840.

Machado, C.J., and Bachevalier, J. (2003). Non-human primate models of childhood psychopathology: The promise and the limitations. *Journal of Child Psychology and Psychiatry*, 44:64–87.

Macchi, G., Rinvik, E. (1976). Thalmo-telencephalic circuits: A neuroanatomical survey. In A. Rémond (Ed.). *Handbook of Electroencephalography and Clinical Neurophysiology (Vol. 2, Pt. A)*. Amsterdam: Elsevier.

Maestro, S., Muratori, F., Cavallaro, M.C., Pei, F., Stern, D., Golse, B., and Palacio-Espasa, F. (2002). Attentional skills during the first 6 months of age in autistic spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry,* 41:1239–1245.

Manes, F., Piven, J., Vrancic, D., Nanclares, V., Plebst, C., and Starkstein, S.E. (1999). An MRI study of the corpus callosum and cerebellum in mentally retarded autistic individuals. *Journal of Neuropsychiatry and Clinical Neurosciences*, 11:470–474.

Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology*, 202:437–470.

Mars, A.E., Mauk, J.E., and Dowrick, P. (1998). Symptoms of pervasive developmental disorders as observed in prediagnostic home videos of infants and toddlers. *The Journal of Pediatrics*, 132:500–504.

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

McCormick, D.A., Clark, G.A., Lavond, D.G., and Thompson, R.F. (1982). Initial localization of the memory trace for a basic form of learning. *Proceedings of the National Academy of Sciences*, 79:2731–2735.

McCormick, D.A., and Thompson, R.F. (1984). Neuronal responses of the rabbit cerebellum during acquisition and performance of a classically conditioned nictating membrane-eyelid response. *Journal of Neuroscience*, 4:2811–2822.

Medin, D.L., Altom, M.W., and Murphy, T.D. (1984). Given versus induced category representation: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10:333–352.

Medin, D.L., and Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, 85:207–238.

Medin, D.L., and Smith, E.E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, 7:241–253.

Miller E.K., Li L., & Desimone R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science,* 254:1377–1379.

Minshew, N.J., Luna, B., and Sweeny, J.A. (1999). Oculomotor evidence for neocortical systems but not cerebellar dysfunction in autism. *Neurology*, 52:917–922.

Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not separate removal of the amygdala and hippocampus. *Nature*, 273:297–298.

Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London B*, 298:85–95.

Mishkin, M. (1993). Cerebral memory circuits. In T.A. Poggio and D.A. Glaser DA (Eds.), *Exploring Brain Functions: Models in Neuroscience*. New York:Wiley and Sons, pp. 113–125.

Mishkin, M. and Aggleton, J. (1981). Multiple functional contributions of the amygdala in the monkey. In: Y. Ben-Ari (Ed.) *The Amygdaloid Complex,* Amsterdam: Elsevier, pp 409-420.

Mishkin, M., and Appenzeller, T. (1987). The anatomy of memory. *Scientific American*, 256:80–89.

Mishkin, M., and Delacour, J. (1975). An analysis of short-term visual memory in the monkey. *Journal of Experimental Psychology: Animal Behavior Processes*, 1:326–334.

Mishkin, M., Ungerleider, L.G., and Macko, K.A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences,* 6:414–417.

Miyata, M., Finch, E.A., Khiroug, L., Hashimoto, K., Hayasaka, S., Oda, S.I., Inouye, M., Takagishi, Y., Augustine, G.J., and Kano, M. (2000). Local calcium release in dendritic spines required for long-term synaptic depression. *Neuron*, 28:233–44.

Mottron, L., Belleville, S., and Menard, E. (1999). Local bias in autistic subjects as evidenced by graphic tasks: Perceptual hierarchization or working memory deficit? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40:743–755.

Mounts, J.R.W. (2000). Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. *Perception and Psychophysics,* 62:969–983.

Muller, R.A., Pierce, K., Ambrose, J.B., Allen, G., and Courchesne, E. (2001). Atypical patterns of cerebral motor activation in autism: A functional magnetic resonance study. *Biological Psychiatry*, 49:665–676.

Mumford, D. (1992). On the computational architecture of the neocortex, II. The role of cortico-cortical loops. *Biological Cybernetics*, 241–251.

Musselwhite, M.J., and Jeffreys, D.A. (1985). The influence of spatial frequency on the reaction times and evoked potentials recorded to grating pattern stimuli. *Vision Research*, 25:1545–1555.

Nosofsky, R.M. (1984). Choice, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10:104–114.

Nosofsky, R.M. (1987). Attention and learning processes in the identification-categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13:87–108.

Nosofsky, R.M. (2000). Exemplar representation without generalization? Comment on Smith and Minda's (2000) "Thirty categorization results in search of a model". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26:1735–1743.

Nosofsky, R.M. and Johnson, M.K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin and Review*, 7(3):375–402.

Nosofsky, R.M., Kruschke, J.K., and McKinley, S.C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18:211–233.

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

Nosofsky, R.M., and Zaki, S.R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28:924–940.

Nowinski, C.V., Minshew, N.J., Luna, B., Takarae, Y, and Sweeney, J.A. (2005). Oculomotor studies of cerebellar function in autism. *Psychiatry Research*, 137: 11-19.

O'Leary, D.D., Borngasser, D.J., Fox, K., and Schlaggar, B.L. (1995). Plasticity in the development of neocortical areas. *Ciba Foundation Symposium,* 193:214–230; discussion 251–257.

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

Olson, S. and Grossberg, S. (1998). A neural network model for the development of simple and complex cell receptive fields within cortical maps of orientation and ocular dominance. *Neural Networks*, 11:189–208.

O'Neill, M. and Jones, R.S.P. (1997). Sensory-perceptual abnormalities in autism: a case for more research? *Journal of Autism and Developmental Disorders*, 27: 283-293.

Optican, L.M., and Robinson, D.A. (1980). Cerebellar-dependent adaptive control of primate saccadic system. *Journal of Neurophysiology*, 44:108–176.

Otsuka, H., Harada, M., Hisaoka, S., and Nishitani, H. (1999). Brain metabolites in the hippocampal-amygdala region and cerebellum in autism: An H-MR spectroscopy study. *Neuroradiology*, 41:517–519.

Otto, T., and Eichenbaum, H. (1992). Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: Evidence for hippocampal processing in recognition memory. *Hippocampus*, 2:323–334.

Paine, R.W., Grossberg, S., and Van Gemmert, A.W.A. (2005). A quantitative evaluation of the AVITEWRITE model of handwriting learning. *Human Movement Science*, 23:837-860.

Page, M. (2000). Connectionist modeling in psychology: a localist manifesto. *Behavioral and Brain Sciences*, 23:443–512.

Palmeri, T.J., Nosofsky, R.M., and McKinley, S.K. (1994). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21:548–568.

Parker, J.L. and Dostrovsky, J.O. (1999). Cortical involvement in the induction, but not expression, of thalamic plasticity. *The Journal of Neuroscience,* 19:8623–8629.

Parker, D.M. and Salzen, E.A. (1977a). Latency changes in the human visual evoked response to sinusoidal gratings. *Vision Research*, 17: 1201–1204.

Parker, D.M. and Salzen, E.A. (1977b). The spatial selectivity of early and late waves within the human visual evoked response. Perception, 6: 85–95.

Parker, D.M., Salzen, E.A., and Lishman, J.R. (1982a). Visual-evoked responses elicited by the onset and offset of sinusoidal gratings: Latency, waveform, and topographic characteristics. *Investigative Ophthalmology and Visual Sciences*, 22: 657–680.

Parker, D.M., Salzen, E.A., and Lishman, J.R. (1982b). The early waves of the visual evoked potential to sinusoidal gratings: Responses to quadrant stimulation as a function of spatial frequency. *Electroencephalography and Clinical Neurophysiology*, 53: 427–435.

Passingham, R.E. (1997). The *Frontal Lobes and Voluntary Action.* Oxford: Oxford University Press.

Pavlov, I.P., (1927) *Conditioned Reflexes*. London: Constable and Company. (Reprinted by Dover Publications, 1960.)

Perner, J., Frith, U., Leslie, A.M., and Leekam, S.R. (1989). Exploration of the autistic child's theory of the mind: Knowledge, belief, and communication. *Child Development*, 60:688–700.

Perrett, D.I., Mistlin, A.J., and Chitty, A.J. (1987). Visual cells responsive to faces. *Trends in Neurosciences*, 10:358–364.

Perret, S.P., Ruiz, B.P., and Mauk, M.D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *Journal of Neuroscience*, 13:1708–1718.

Pickles, A., Starr, E., Kazak, S., Bolton, P., Papanikolaou, K., Bailey, A., Goodman, R., and Rutter, M. (2000). Variable expression of the autism broader phenotype: Findings from extended pedigrees. *Journal of Child Psychology and Psychiatry,* 41:491–502.

Pierce, K., Haist, F., Sedaghat, F., and Courchesne, E. (2004). The brain response to personally familiar faces in autism: findings of fusiform activity and beyond. *Brain,* 127:2703–2716.

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

Piven, J., Saliba, K., Bailey, J., and Arndt, S. (1997). An MRI study of autism: the cerebellum revisited. *Neurology*, 49:546–51.

Plant, G.T., Zimmern, R.L., and Durden, K. (1983). Transient visually evoked potentials to the pattern reversal and onset of sinusoidal gratings. *Electroencephalography and Clinical Neurophysiology*, 56: 147–158.

Pollen, D.A. (1999). On the neural correlates of visual perception. *Cerebral Cortex*, 9:4–19.

Posner, M.I., and Keele, S.W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, 83:304–308.

Raizada, R.D.S., and Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex*, 13:100–113.

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. *Nature Neuroscience*, 2:79–87.

Receveur, C., Lenoir, P., Desombre, H., Barthelmy, C., and Malvy, J. (2005). Interaction and imitation deficits from infancy to 4 years of age in children with autism: a pilot study based on videotapes. *Autism*, 9: 69-82.

Reynolds, J.H. and Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27:611–647.

Reynolds, J., Chelazzi, L. and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of Neuroscience*, 19:1736–1753.

Rinehart, N.J., Bradshaw, J.L., Brereton, A.V., and Tonge, B.J. (2001). Movement preparation in high-functioning autism and Asperger disorder: A serial choice reaction time involving motor reprogramming. *Journal of Autism and Developmental Disorders,* 31:79–88.

Ring, H.A., Baron-Cohen, S., Wheelwright, S., Williams, S.C., Brammer, M., Andrew, C., and Bullmore, E.T. (1999). Cerbral correlates of preserved cognitive skills in autism: A functional MRI study of embedded figures task performance. *Brain*, 122:1305–1315.

Risch, N., Spiker, D., Lotspeich, L., Nouri, N., Hinds, D., Hallmayer, J., Kalaydjieva, L., McCague, P., Dimiceli, S., Pitts, T., Nguyen, L., Yang, J., Harper, C., Thorpe, D., Vermeer, S., Young, H., Hebert, J., Lin, A., Ferguson, J., Chiotti, C., Wiese-Slater, S., Rogers, T., Salmon, B., Nicholas, P., and Myers, R.M. (1999). A genomic screen of autism: Evidence for a multilocus etiology. *American Journal of Human Genetics,* 65:493–507.

Ritvo, E.R., Freeman, B.J., Scheibel, A.B., Duong, T., Robinson, H., Guthrie, D., and Ritvo, A. (1986). Lower Purkinje cell count in the cerebella of four autistic subjects: Initial findings of the UCLA-NSAC Autopsy Research Report. *American Journal of Psychiatry*, 143:862–866.

Roelfsema, P.R., Lamme, V.A.F. and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395:376–381.

Rolls, E.T. (1998). The orbitofrontal cortex. In A.C. Roberts, T.W. Robbins, L. Weiskrantz , (Eds.); *The Prefrontal Cortex: Executive and Cognitive Functions*, Oxford: Oxford University Press, pp 67-86.

Rutter, M. (2000). Genetic studies of autism: From the 1970s into the millennium. *Journal of Abnormal Child Psychology,* 28:3–14.

Ryle, G. (1949). *The Concept of Mind*. New York:Hutchinson's University Press.

Saitoh, O., and Courchesne, E. (1998). Magnetic resonance imaging study of the brain in autism. *Psychiatry and Clinical Neurosciences*, 52:19–22.

Schlaggar, B.L., Fox, K., and O'Leary, D.D. (1993). Postsynaptic control of plasticity in developing somatosensory cortex. *Nature*, 364:623–626.

Schoenbaum, G., Setlow, B., Saddoris, M.P., and Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron*, 39:855–867.

Schwartz, E.L., Desimone, R., Albright, T., and Gross, C.G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences*, 80:5776–5778.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesion. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20:11–21.

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

Sigman, M., Yirmiya, N., and Capps, L. (1995). Social and cognitive understanding in high-functioning children with autism. In Schpler, E., and Mesibov, G.B. (Eds.), *Learning and Cognition in Auitsm.* New York:Plenum Press, pp. 159–176.

Sillito, A.M., Jones, H.E., Gerstein, G.L. and West, D.C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, 369:479–482.

Singer, W. (1983). Neuronal activity as a shaping factor in the self-organization of neuron assemblies. In: *Synergetics of the brain* (E. Basar, H. Flohr, H. Haken, and A.J. Mandell, Eds.). New York: Springer-Verlag.

Skrandies, W. (1984). Scalp potential fields evoked by grating stimuli: Effects of spatial frequency and orientation. *Electroencephalography and Clinical Neurophysiology*, 58: 325–332.

Smith, A.T., Singh, K.D. and Greenlee, M.W. (2000). Attentional suppression of activity in the human visual cortex. *Neuroreport,* 11:271–277.

Smith, J.D., and Minda, J.P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 24:1411–1436.

Smith, J.D., and Minda, J.P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 26:3–27.

Smith, J.D., Murray, M.J., and Minda, J.P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 23:659–680.

Smith, M.C. (1968). CS-US interval and US intensity in classical conditioning of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*, 3:678–687.

Sokol, D.K., and Edwards-Brown, M. (2004). Neuroimaging in autistic spectrum disorder (ASD). *Journal of Neuroimaging,* 14:8–15.

Somers, D.C., Dale, A.M., Seiffert, A.E., and Tootell, R.B.H. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex, *Proceedings of the National Academy of Science, USA*, 96:1663–1668.

Sparks, B.F., Friedman, S.D., Shaw, D.W., Aylward, E.H., Echelard, D., Artru, A.A., Maravilla, K.R., Giedd, J.N., Munson, J., Dawson, G., and Dager, S.R. (2002). Brain structural abnormalities in young children with autism spectrum disorder. *Neurology,* 59:184–92.

Spitzer, H., Desimone, R., and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240:338–340.

Squire, L.R., and Cohen, N.J. (1984). Human memory and amnesia. In G. Lynch, J. McGaugh, and N.M. Weinberger (Eds.), *Neurobiology of Learning and Memory*. New York:Guilford Press, pp. 3–64.

Squire, L.R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253:1380–1386.

Staddon, J.E.R. (1983). *Adaptive Behavior and Learning*. New York:Cambridge University Press.

Steinman, B.A., Steinman, S.B. and Lehmkuhle, S. (1995). Visual attention mechanisms show a canter-surround organization. *Vision Research*, 35:1859–1869.

Sturm, H., Fernell, E., and Gillberg, C. (2004). Autism spectrum disorders in children with normal intellectual levels: associated impairments and subgroups. *Developmental Medicine and Child Neurology,* 46:444–447.

Sweeten, T.L., Posey, D.J., Shekar, A., and McDougle, C.J. (2002). The amygdala and related structures in the pathophysiology of autism. *Pharmacology, Biochemistry, and Behavior*, 71(3):449–455.

Szelag, E., Kowalska, J., Galkowski, T., and Poppel, E. (2004). Temporal processing deficits in high-functioning children with autism. *British Journal of Psychiatry.* 95:269–282.

Tager-Flusberg, H., Joseph, R., and Folstein, S. (2001). Current Directions in Research on Autism. *MRDD Research Reviews.* 7:21-29.

Takarae, Y., Minshew, N.J., Luna, B., and Sweeney, J.A. (2004). Oculomotor abnormalities parallel cerebellar histopathology in autism. *Journal of Neurology, Neurosurgery, and Psychiatry,* 75(9):1359–1361.

Takechi, H., Eilers, J., and Konnerth, A. (1998). A new class of synaptic response involving calcium release in dendritic spines. *Nature*, 396:757–760.

Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., and Maurer, R.G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences USA*, 95:13982–13987.

Temereanca, S. and Simons, D.J. (2001). Topographic specificity in the functional effects of corticofugal feedback in the whisker/barrel system. *Society for Neuroscience Abstracts,* 393.6.

Thompson, R.F. (1988). The neural basis of basic associative learning of discrete behavioral responses. *Trends in Neurosciences*, 11:152–155.

Thompson, R.F., Barchas, J.D., Clark, G.A., Donegan, N., Kettner, R.E., Lavond, D.G., Madden, J., Mauk, M.D., and McCormick, D.A. (1984). Neuronal substrates of associative learning in the mammalian brain. In D.L. Aldon and J. Farley (Eds.), *Primary Neural Substrates of Learning and Behavioral Change*. New York:Cambridge University Press, pp. 71–99.

Thompson, R.F., Clark, G.A., Donegan, N.H., Lavond, G.A., Lincoln, D.G., Maddon, J., Mamounas, L.A., Mauk, M.D., and McCormick, D.A. (1987). Neuronal substrates of discrete, defensive conditioned reflexes, conditioned fear states, and their interactions in the rabbit. In I. Gormenzano, W.F. Prokasy and R.F. Thompson (Eds.), *Classical Conditioning*, *Third Edition.* Hillsdale, NJ: Erlbaum Associates, pp. 371–399.

Townsend, J., Courchesne, E., Covington, J., Westerfield, M., Harris, N.S., Lyden, P., Lowry, T.P., and Press, G.A. (1999). Spatial attention deficits in patients with acquired or developmental cerebellar abnormality. *Journal of Neuroscence,* 19:5632–5643.

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. *Neural Computation*, 10:821–835.

Tsumoto, T., Creutzfeldt, O.D., Legéndy, C.F. (1978). Functional organization of the corticofugal system from visual cortex to lateral geniculate nucleus in the cat. *Experimental Brain Research*, 32:345-364.

Tuchman, R., and Rapin, I. (1997). Regression in pervasive developmental disorders: seizures and epileptiform electroencephalogram correlates. *Pediatrics,* 99:560–566.

Ungerleider, L.G. and Mishkin, .M (1982). Two cortical visual systems: Separation of appearance and location of objects. In D.L. Ingle, M.A. Goodale and R.J.W. Mansfield (Eds.) *Analysis of Visual Behavior*, Cambridge: MIT Press, 549–586.

Vanduffel, W., Tootell, R.B. and Orban, G.A. (2000). Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. *Cerebral Cortex,* 10:109–126.

Vassilev, A., Manahilov, V., and Mitov, D. (1983). Spatial frequency and pattern onset-offset response. *Vision Research,* 23: 1417–1422.

Vassilev, A. and Strashimirov, D. (1979). On the latency of human visually evoked response to sinusoidal gratings. *Vision Research*, 19: 843–846.

Vilensky, J.A., Damasio, A.R., and Maurer, R.G. (1981). Gait disturbances in patients with autistic behavior: A preliminary study. *Archives of Neurology*, 38:646–649.

Vinogradova, O.S., (2001). Hippocampus as comparator: Role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus,* 11:578–598.

Volkmar, F.R.and Pauls, D. (2003). Autism. *Lancet,* 362:1133-1141.

Von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100.

Wing, L. (1997). The autistic spectrum. *Lancet.* 350:1761–1766.

Werner, E., Dawson, G., Osterling, J., and Dinno, J. (2000). Recognition of autistic spectrum disorders before 1 year of age: A retrospective study based on home videotapes. *Journal of Autism and Developmental Disorders*, 30:157–162.

Williamson, S.J., Kaufman, I., and Brenner, D. (1978). Latency of the neuromagnetic response of the human visual cortex. *Vision Research*, 18: 107–110.

Wilson, F.A.W., Scalaidhem, O., Goldman-Rakic, P.S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex.. *Science,* 260:1955-1958.

Yazdanbakhsh, A. and Grossberg, S. (2004). Fast synchronization of perceptual grouping in laminar visual cortical circuits. *Neural Networks*, 17:707–718.

Yoshida, Y., and Uchiyama, T. (2004). The clinical necessity for assessing Attention Deficit/Hyperactivity Disorder (AD/HD) symptoms in children with high-functioning Pervasive Developmental Disorder (PDD). *European Child and Adolescent Psychiatry.* 13:307–314.

Zaki, S.R., Nososfsky, R.M., Jessup, N.M. and Unverzagt, F.W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society*, 9:394–406.

Zhang, Y., Suga, N., and Yan, J. (1997). Corticofugal modulation of frequency processing in bat auditory system. Nature, 387:900 –903.