Self-Organization of an Artificial Neural Network Subjected to Attention Shift Impairments and Familiarity Preference, Characteristics Studied in Autism

Lennart Gustafsson[1] and Andrew P. Papliński[2,3]

Autism is a developmental disorder with possibly multiple pathophysiologies. It has been theorized that cortical feature maps in individuals with autism are inadequate for forming abstract codes and representations. Cortical feature maps make it possible to classify stimuli, such as phonemes of speech, disregarding incidental detail. Hierarchies of such maps are instrumental in creating abstract codes and representations of objects and events. Self-Organizing Maps (SOMs) are artificial neural networks that offer insights into the development of cortical feature maps.

Attentional impairment is prevalent in autism, but whether it is caused by attention-shift impairment or strong familiarity preference or negative response to novelty is a matter of debate. We model attention shift during self-organization by presenting a SOM with stimuli from two sources in four different modes, namely, novelty seeking (regarded as normal learning), attention-shift impairment (shifts are made with a low probability), familiarity preference (shifts made with a lower probability to the source that is the less familiar to the SOM of the two sources), and familiarity preference in conjunction with attention-shift impairment.

The resulting feature maps from learning with novelty seeking and with attention-shift impairment are much the same except that learning with attention-shift impairment often yields maps with a somewhat better discrimination capacity than learning with novelty seeking. In contrast, the resulting maps from learning with strong familiarity preference are adapted to one of the sources at the expense of the other, and if one of the sources has a set of stimuli with smaller variability, the resulting maps are adapted to stimuli from that source. When familiarity preference is less pronounced, the resulting maps may become normal or fully restricted to one of the sources, and in that case, always the source with smaller variability if such a source is present. Such learning, in a system with many different maps, will result in very uneven capacities.

Learning with familiarity preference in conjunction with attention-shift impairment surprisingly has higher probability for the development of normal maps than learning with familiarity preference alone.

KEY WORDS: Autism; attention; neural networks; self-organizing maps.

INTRODUCTION—AUTISM, ATTENTIONAL IMPAIRMENT, AND FAMILIARITY PREFERENCE

Autism is a developmental disorder first described by Kanner (Kanner, 1943), and Asperger (Asperger, 1944). At present, diagnostic criteria according to the *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (DSM-IV; American Psychiatric Association, 1994) are grouped into three main categories:

[1] Luleå University of Technology, S-971 87 Luleå, Sweden.
[2] Computer Science and Software Engineering, Monash University, Vic 3800, Australia.
[3] Correspondence should be addressed to Andrew P. Papliński, Computer Science and Software Engineering, Monash University, Vic 3800, Australia; e-mail: app@csse.monash.edu.au

impairments in social interaction, impairments in verbal and nonverbal communication, and restricted repertoire of activities and interests.

The diagnostic criteria are behavior-based, but a number of biological abnormalities have been connected with autism. For an introduction, see Gillberg and Coleman (2000).

There is a general agreement that attentional impairment is commonly seen in autism. This impairment includes joint attention and attention shifts. There are, however, different opinions as to the precise nature of the attentional impairment.

A large body of research on attentional impairment has been presented by Courchesne and coworkers (see Courchesne *et al.,* 1994b, 1994a; Courchesne, Akshoomoff, Townsend, & Saitoh, 1995) and Townsend *et al.* (Townsend, Harris, & Courchesne, 1996; Townsend *et al.,* 1999).

The importance of cerebellar damage (loss of Purkinje neurons, documented in many cases of autism and hypothesized by Courchesne to take place during infancy; Courchesne, 2002) in impairing the ability to shift attention is stressed in this research. Courchesne *et al.* argue that impairment in shifting attention will cause autism because memories of events will be incomplete and fragmented if attention-shift impairments preclude a perception of some aspects of the events. They also hypothesize that this will lead to the development of other autistic characteristics, among them obsessive insistence on sameness (Courchesne *et al.,* 1994b, pp. 104–105).

Other researchers present results that indicate that there is no general deficit in the ability to shift attention in children with autism, but that attention-shift impairment is manifested particularly when social stimuli are present. Dawson, Meltzoff, Osterling, Rinaldi, and Brown (1998) report that children with autism orient to stimuli from inanimate objects such as rattles almost as well as normal children, whereas stimuli such as hand clapping resulted in many orienting errors by children with autism. Pascualvaca, Fantie, Papageorgiou, and Mirsky (1998) found that children with autism had difficulty in achieving a task which included shifting mental set. Interestingly, they behaved like normal children initially, that is, they could shift set, but after having chosen some problem-solving strategy, they could not change it and did not benefit from feedback from the examiner. Minshew, Luna, and Sweeney (1999) have studied saccadic eye movements in individuals with autism and found that such movements, when under cerebellar control, are normal, but under neocortical influence, they are not. Minshew *et al.* state

that deficits in "elementary attentional and sensorimotor systems" (p. 917) were not demonstrated.

Dawson, Pascualvaca, and other researchers hypothesize that novelty itself is disagreeable to children with autism (social stimuli are assumed to be connected with novelty more often than nonsocial stimuli) and that novelty avoidance will cause attention impairments and, as a consequence cause autistic characteristics. Kootz, Marinelli, & Cohen (1982) note that

> Rather than responding to novelty with orientation, observation, and exploration, the autistic child often responds with avoidance, thus preventing the development of new schemas and subsequent familiarization. (p. 192)

The insight that children with autism dislike novelty, or at least prefer familiarity, is not new. Kanner (1943) originally described the cases he studied as possessing two cardinal features, one of them being obsessive insistence on sameness.

It is the purpose of this article to use a simple neural network model to study developmental differences that emerge between cases in which the neural network is subjected to an unconditional attention-shifting impairment and cases in which familiarity preference causes attention shifting to cease.

ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANNs) consist of synapses with adjustable connection-strength parameters (weights) and signal-aggregating nodes representing neurons. During the learning process, weight vectors are modified according to specific features of the presynaptic signals. ANNs are used for information processing after periods of learning from typical examples and for simulation of biological neural networks. For a general introduction to artificial neural networks, see, for example, Haykin (1999). Of particular importance are the self-organizing maps (SOMs) developed by Kohonen (2001), which produce topographical feature maps that may represent feature maps in sensory cortices. There are very convincing demonstrations of the correspondence between self-organizing maps and measured cortical feature maps of animals (Ritter, Martinetz, & Schulten, 1992). Further arguments for such correspondence have been presented by Spitzer (1995). For an introduction to cortical maps, see, for example, Kandel, Schwartz, and Jessel (2000), and for a treatment of the development of cortical maps, including modeling aspects of cortical maps, see Price and Willshaw (2000).

Theories on causes of autism, based on properties of artificial neural networks, have been presented by Cohen (1994, 1998), Gustafsson (1997), and McClelland (2000). An analysis based on artificial neural networks simulations of the emergence of developmental disorders, including, but not restricted to, autism, has been presented by Oliver, Johnson, and Pennington (2000). An artificial neural network is subjected to a learning process to enable it to detect and categorize stimuli presented during the learning process. The learning process of a SOM simulates the development and fine-tuning of a cortical feature map. Through hierarchies of such cortical feature maps in sensory cortices, ever more abstract representations of objects and events enable the coding of experiences without involving masses of details. It was found by Hermelin (1978) that autistic children are impaired in their capacity to recode information from sensory to abstract codes, making it difficult for them to see what normal individuals regard as salient features of a situation (see, e.g., Happé, 1991). A learning process for an artificial neural network may result in an inadequate feature map such that correct classification of stimuli, ignoring incident detail, is not well accomplished. In the theories presented by Cohen and Gustafsson, this is seen as modeling the impairments in forming more abstract codes and representations, evident in individuals with autism.

In this article, the artificial neural networks themselves are assumed to be adequate for successful self-organization. The purpose of this article is to examine how attention-shift impairment and familiarity preference influence the self-organization of an artificial neural network and to discuss the characteristics of the resulting maps. It will be shown that some, but not all,

of these maps exhibit characteristics that make them inadequate for categorization of stimuli; that is, some classes of stimuli presented during the learning process will not be categorized. A comparison will be made with maps resulting from self-organization when novelty seeking is present.

SIMULATIONS—METHODS

In the simulations presented below, two sources generate two-dimensional stimuli. The stimuli provided by the sources are represented in the figures (e.g., see Fig. 1) by "0" and "+," respectively. In most simulations we use three groups of data, each containing 10 stimuli (i.e., points in a two-dimensional space) in each source, or 60 points altogether (in the software model, all parameters can be easily modified). The sources can be thought of as producing, for example, two dialects of a very limited protolanguage, each with three protophonemes. We can imagine that a source is a representation of a parent of a child pronouncing three phonemes in 10 slightly different ways. The parallel is far from perfect, but it might be helpful for a conceptual understanding of the simulations.

Real sensory stimuli, like the phonemes of speech, are of course larger in number and dimension. Simulations with sources that provide few classes of this low dimensionality have been presented because a complete visualization of the data and the resulting neuronal weights can easily be achieved in two-dimensional space. Hence, the simulations that follow are, compared to any test situation in which humans are involved, highly simplified—they are unimodal, that is, only one
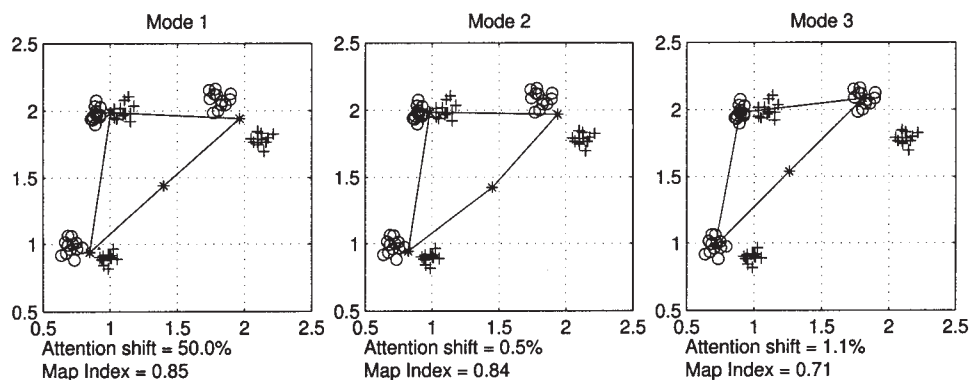
Fig. 1. Learning in a canonical SOM with nodes in a 2 × 2 mesh. A pair of *x*, *y* axes specify the input and weight (feature) space; that is, the strength of the stimuli and the values of weights associated with the neuronal nodes. The resulting maps are from self-organization in mode 1, mode 2, and mode 3, respectively. The relative frequency of attention shifts during learning is shown below each map. Map Index is explained later in the text.

sensory system is involved; the stimuli are very simple, that is, each can be described by a two-dimensional number; and they are very few. This allows for a high degree of control of the experiments and for visualization of the results.

One of the important parameters of data is its variability (or spread) that we measure by the trace of the respective covariance matrices. In some simulations we move one cluster of data to reduce the variability for that specific source.

Important also to our simulations is the mechanism that shifts attention between two sources of stimuli. New stimuli are made available by these sources at random, although by only one source at a time. The artificial neural network is engaged in a particular source when the output from that source constitutes the input to the artificial neural network. Shift of engagement to the alternate source is determined in four different modes in the self-organization of the artificial neural network.

Mode 1, Novelty Seeking

Attention is shifted to the alternate source if that source presents the next new stimulus. Mode 1 is regarded as the "normal" mode of learning.

Mode 2, Attention-Shift Impairment

Attention is shifted to the alternate source with a low probability, in the simulations chosen as 0.01, if that source presents the next new stimulus. The results obtained are robust against changes of the probability for shifting. Other values of the probability, both much smaller and larger, have been tested with no observed differences in the results. In the extreme case, when the probability is set to 0, then clearly the map will self-organize to adapt to stimuli from one source only.

Mode 3, Familiarity Preference

Attention is shifted to the alternate source if that source presents the next new stimulus, but conditionally, depending on the map's familiarity with the source. The map familiarity to a particular source is measured by the time-averaged value of the distance between map nodes and the stimuli. When both sources are unfamiliar to the map that is, in the beginning stage of the self-organization, attention is shifted to the alternate source if that source presents the next stimulus. As the map develops some familiarity with the sources; that is, the node weights begin to resemble the data, attention is shifted with a higher probability to the source that is most familiar to the map. The degree of familiarity preference will be measured by this probability of attention shift.

Mode 3 is intended to model a self-organization that has the same attention shift characteristics as those reported by Pascualvaca *et al.* (1998).

Mode 4, Familiarity Preference in Conjunction with Attention-Shift Impairment

Attention is shifted to the alternate source if that source presents the next new stimulus according to the conditions stated in mode 3, but the probability for a shift is conjoined with a low probability, chosen as 0.01 in the simulations.

The neuronal nodes are also, as is the most common case, organized in a two-dimensional rectangular mesh, and the neurons in these simulations have two weights each; that is, the weight vector and the source data are both two-dimensional. The node weight vectors are represented by "*." It is expected that in a "normal" map, neuronal nodes will be moved to the centers of the data clusters. This feature will be measured by the Map Index explained below.

The following technical comments are instrumental to the map formation algorithm. The initial values of the weights were selected randomly in a small square around the center of stimuli (see Fig. 2). This is a very noncritical starting point for the maps to develop. Other selections affect slightly the map formation speed. The learning rate, η_n, is initially chosen as $\eta_0 = 0.1$, then decreased after each epoch n according to the formula $\eta_n = \eta_0/(1 + 0.01n)$. The learning rate is always decreased as learning proceeds, which makes an initial map formation and its fine tuning possible. The choice of learning rate is not critical for the final maps, but has an influence on the process of self-organization. The learning rate must be chosen that allows both the initial map formation and its fine tuning. The small final value of the learning rate secures the stability of the map formation algorithm.

The number of epochs in each simulation was chosen as $2^{10} = 1024$ to ensure stable final maps. We tried as many as $2^{12} = 4096$ epochs, but because no changes of the maps were observed after 1024 simulations, we chose the smaller number of epochs, which has an obvious speed advantage.

SIMULATIONS—RESULTS

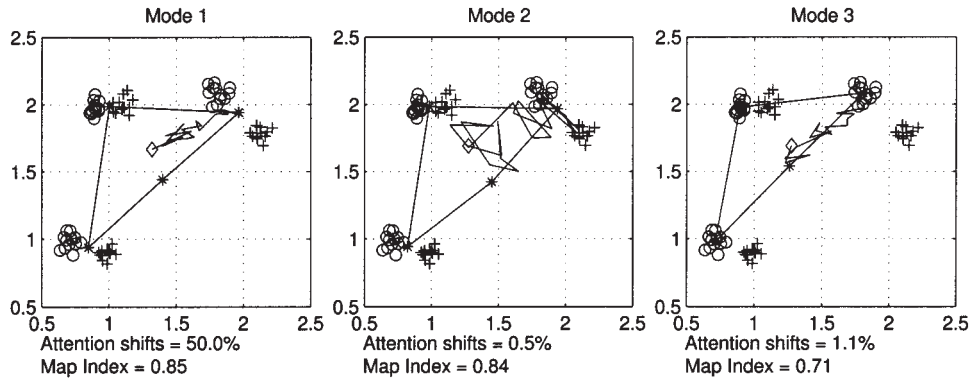In the first set of simulations, a canonical SOM (the smallest possible ANN that can successfully learn

Fig. 2. Simplified trajectories of selected nodes during self-organization. The initial position of the relevant weight vector is marked with "◇."

to classify the protophonemes but not distinguish their dialectal forms) with neuronal nodes organized in a 2×2 rectangular mesh was used (two-dimensional SOMs are generally arranged with nodes in a rectangular or hexagonal pattern). The resulting maps from one simulation of learning, that is, self-organization, in modes 1, 2, and 3 are shown in Figure 1. In normal learning (mode 1), because of the fact that the total numbers of stimuli in each source are equal, the frequency of the attention shift will be close to 50%. In learning with attention-shift impairment (mode 2), the frequency of the attention shift will be reduced by the probability of the attention shift and therefore will be close to $0.01 \times 50\% = 0.5\%$. As can be seen normal learning with novelty seeking, and learning with attention-shift impairment, results in very similar maps—the node weights assume values that are the mean values of each class. Learning with familiarity preference, however, results in a map in which the nodes have adapted with a preference for the stimuli from one source. This result has been repeated, although the preferred source has altered, in many hundreds of simulations.

To compare the resulting maps, a Map Index has been introduced. The Map Index is a value between 0 and 1 that is formed from a normalized squared distance between the mean value of each subclass (or protophoneme of a dialect) and the closest neuronal node. The Map Index thus has its highest value when there is a node for each subclass; that is, when not only each protophoneme is identified but also the dialect of each protophoneme. When the map consists of only four nodes this cannot be attained, and then the best possible result is to have one node for the mean value of each class so that each protophoneme is identified, even though the dialects cannot be identified. No credit is

given in the Map Index for a node coinciding with a particular stimulus; the merit of the map consists of it adapting to the means of each subclass, or class, when that is what the map's capacity allows, and not in the details represented by the data of each individual stimulus.

The statistics of the Map Index from 100 simulations is shown in Table I. In each of the 100 simulations, the stimuli are different but have the same mean values. Likewise, the initial values of the nodal weights are chosen with a small random variation around the same mean values. Each time, the order of presentation of data points is randomly changed.

It might at first seem surprising that the resulting maps from learning with attention-shift impairment are almost the same as those resulting from learning with novelty seeking, but the explanation is straightforward— if the nodes were more adapted to the mean values of the subclasses of one of the sources, then learning from

Table I. Statistics of the Map Index from 100 Simulations of Learning in a Network, as in Figure 1

	Mode 1	Mode 2	Mode 3
Mean	0.846	0.836	0.692
SD	0.001	0.012	0.015
Minimum	0.844	0.795	0.651
Maximum	0.847	0.847	0.727

99.9% Confidence Intervals for the Differences between Mode Means

Modes 1–2	Modes 1–3	Modes 2–3
0.01 ± 0.006	0.154 ± 0.005	0.144 ± 0.006

Note: The results of the significance test for the differences between the mode means with the 99.9% confidence are also included.

exemplars of the other source would result in greater node weight adjustments, pulling the node toward the mean value of the subclasses of both sources. In learning with familiarity preference, this mechanism is not present—when the nodes have adjusted more to the subclasses of one of the sources, exemplars from the other source will not cause learning but will be ignored.

Even though the resulting map is the same in modes 1 and 2, the self-organization process is markedly different. After some initial "wandering about," the nodes converge rather undramatically toward their final positions in mode 1 but oscillate wildly before settling down in mode 2. The trajectories of one of the nodes during self-organization in all three modes are shown in Figure 2. The trajectories do not include all intermediate positions of the nodes during self-organization but illustrate the qualitative difference between the self-organizing processes in the different modes.

In the second set of simulations, to test further the interaction between the representational resources of the network and the distributional structure of the input stimuli, we use a SOM with nodes organized in a 3×3 rectangular mesh. Such a bigger SOM has the capacity to discriminate between the dialectal protophoneme forms. In terms of our example domain, this is like the child having more available categories into which to slot his or her developing knowledge of incoming vocabulary items. The maps obtained in this set of simulations are shown in Figure 3. Learning in mode 1 results in maps that, in some cases, have nodes adapted to the means of the subclasses from both sources combined, and in other cases, adapted to the means of the subclasses from each source. Learning in mode 2 results in nodes becoming adapted to the means of the
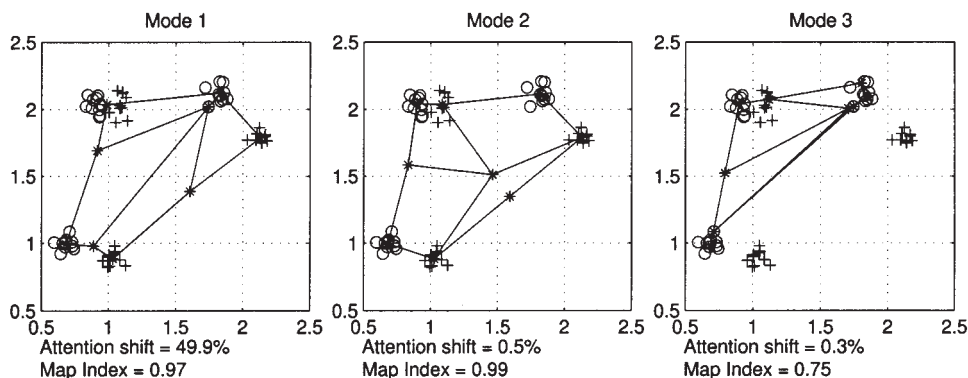
subclasses from each source. Learning in mode 3 resulted in maps that give preference to one source, and most nodes will adapt to subclasses from that source, but nodes may also adapt to one or more subclasses of the other source. Notice that four nodes have adapted to one subclass in mode 3. This means that this map has adapted to the structure of one of the subclasses.

It should be noted that learning in mode 2, that is, with attention-shift impairment, statistically results in a higher Map Index than learning in mode 1, that is, normal learning (see Table II). The chosen criterion for the Map Index rewards the capacity of discrimination, and attention-shift impairment apparently brings some benefits in this respect.

In the third and fourth sets of simulations, one source is reduced to provide only two classes of stimuli, with one of the classes having 20 exemplars, thus reducing the data variability of this source. Referring to our example domain, we may say that now one

Table II. Statistics of the Map Index from 100 Simulations of Learning in a Network, as in Figure 3

	Mode 1	Mode 2	Mode 3
Mean	0.988	0.997	0.755
SD	0.015	0.002	0.037
Minimum	0.948	0.979	0.680
Maximum	0.999	0.999	0.847

99.9% Confidence Intervals for the Differences between Mode Means

Modes 1–2	Modes 1–3	Modes 2–3
-0.009 ± 0.005	0.233 ± 0.013	0.222 ± 0.013

Fig. 3. Learning with in a 3×3 mesh. The resulting maps are from self-organization in mode 1, mode 2, and mode 3, respectively.
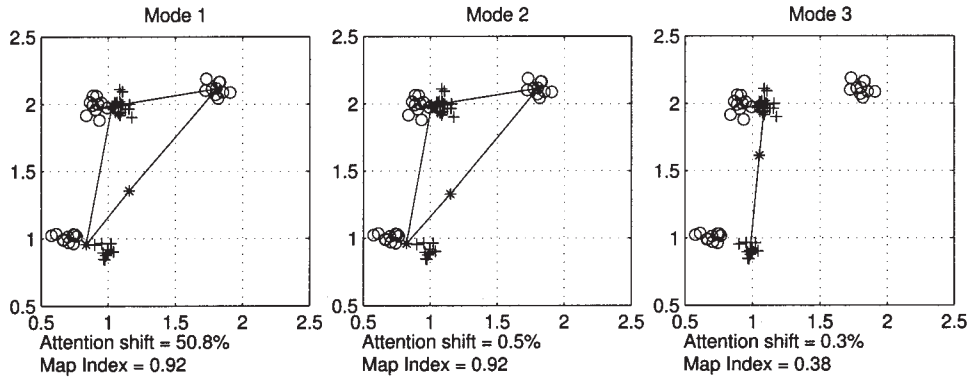
Fig. 4. Learning in a canonical SOM with nodes in a 2 × 2 mesh. The resulting maps are from self-organization in mode 1, mode 2, and mode 3, respectively.

parent pronounces three phonemes but the other parent only two phonemes.

In the third set a canonical 2 × 2 SOM with the same degree of familiarity preference as in the first and second sets of simulations is used; in the fourth set a canonical 2 × 2 SOM with a considerably weaker familiarity preference is used.

The results of the third set of simulations were, with small variations, those shown in Figure 4. As before, the resulting maps from learning in mode 1 and 2 are similar and adequately cover all classes of the stimuli. The resulting map from learning in mode 3 shows that the source with reduced data variability dominates the development of learning, leaving one subclass of stimuli from the full source without any detector node. The statistics given in Table III show that the simulation results presented in Figure 4 are typical. Again there is a straightforward explanation

Table III. Statistics of the Map Index from 100 Simulations of Learning in a Network, as in Figure 4

	Mode 1	Mode 2	Mode 3
Mean	0.926	0.930	0.389
SD	0.001	0.023	0.014
Minimum	0.922	0.873	0.346
Maximum	0.927	0.974	0.435

99.9% Confidence Intervals for the Differences between Mode Means

Modes 1–2	Modes 1–3	Modes 2–3
−0.004 ± 0.008	0.537 ± 0.005	0.541 ± 0.010

Note: The difference between the means of the modes 1 and 2 is not significant.

for this result—during the early stage of self-organization, all the nodes are pulled toward the area of densest data, and the source with the smallest variability will have more of its exemplars there than the source with higher variability (and thus more spread-out data). Thus, the map will start to become more adapted to the source with less variability, that is, its familiarity with that source will grow more than its familiarity with the other source, and according to the mechanism of familiarity preference, the less familiar source will eventually cease to cause attention shifts, whereupon the map will fully adapt to the source with less variability.

In the fourth set of simulations, learning in mode 3 only is investigated. In this case, a weaker familiarity preference than in the third set of simulations has been applied. The resulting maps are shown in Figure 5. Learning in mode 3 here produces very different results in different simulations, even with very small differences in stimuli and initial weights. The stimuli are presented in a random order, and the order is therefore also different in different simulations. The small simulation differences yield maps of such extremes as shown in Figure 5 and in the third column of Figure 4. The Map Index (the third plot in Fig. 5) is strongly dependent on the relative frequency of attention shifts—50% attention shifts invariably yield a high Map Index, whereas a low percentage of attention shifts invariably yields a low Map Index.

Learning in mode 4, that is, with attention-shift impairment in conjunction with familiarity preference, yields the same kind of results as learning in mode 3, but the probability for a normal map is surprisingly considerably larger. This is illustrated in Figure 6 for the same conditions as those documented in Figure 5, augmented with an attention-shift test according to Mode 2.
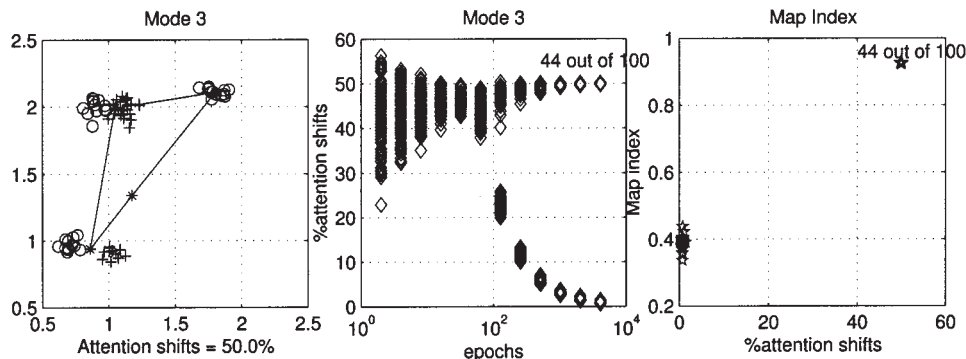
Fig. 5. Learning in a SOM with 2 × 2 nodes. The first column illustrates a successfully self-organized map from learning in mode 3. Notice the number of attention shifts during learning shown below the map. The second column shows the relative number of attention shifts as self-organization proceeds. A total of 100 simulations are represented in this diagram. The third column shows the Map Index of the resulting maps and the corresponding relative number of attention shifts during self-organization.

ROBUSTNESS

In the simulations presented, the stimuli, the initial weights, and the order of presentation of stimuli have all been varied between simulations. The mean values of the stimuli have, however, remained the same. In this section we briefly present results when one class has been added to each source.

In Figure 7 it is seen that the characteristics of learning in modes 1, 2, and 3 are the same as before. The statistics from 100 simulations, not shown here, verify this and show that the simulation results in Figure 7 are typical.

In Figure 8 one of the classes of the "+"-marked source has been moved. The self-organization in modes 1 and 2 show the same characteristics as before. In mode 3, the nodes are adapted to the data of the "+"-marked source, rather than the data of the "o"-marked source as in Figure 7. It is here important

to note that the variability of the data is smallest for the "o"-marked source in Figure 7 and for the "+"-marked source in Figure 8. As before, the self-organization in mode 3 results in maps that are adapted to the source with the smallest variability.

Higher-dimensional data are not considered in this article. However, the explanations given for the obtained results do not rely on the specific dimensionality of the data. Ongoing simulations with stimuli of much higher dimensionality than two are expected to confirm the general validity of the results presented in this article.

DISCUSSION

Our results show that learning with familiarity preference results in maps with characteristics such that the stimuli of one source will be learned precisely at the expense of the other. If one source has stimuli of
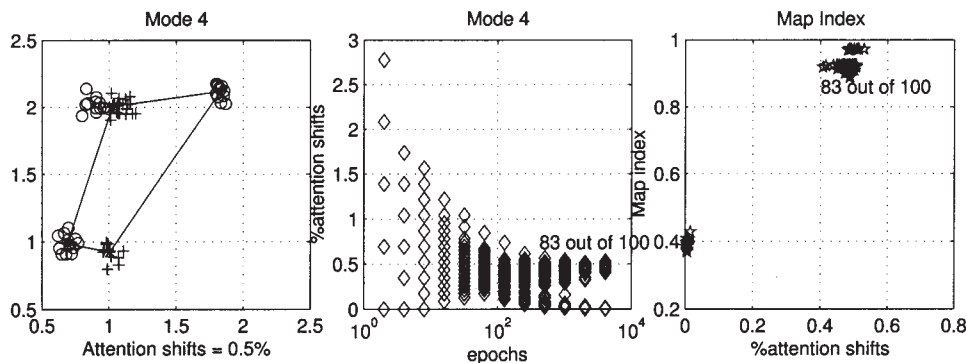
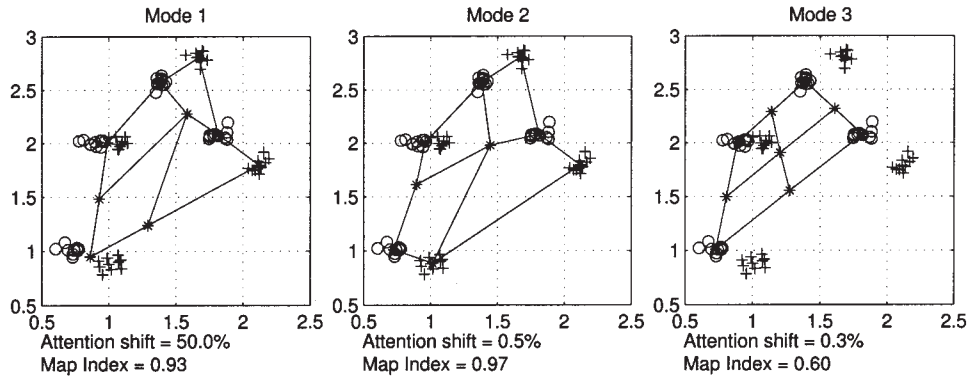

Fig. 6. Learning in a SOM with 2 × 2 nodes in mode 4.

Fig. 7. Learning data set 2 in a 3 × 3 SOM.

smaller variability, it will in many cases dominate the resulting map to the extent that some stimuli from the other source will not be detected. Learning in great detail in a narrow field is a known autistic characteristic (reflected also in the diagnostic criteria), and in our simulations this corresponds to a map adapting to the source with the smallest variability. Learning with attention-shift impairment consistently results, after rather chaotic-seeming learning processes, in normal maps, often with a somewhat better capacity for discrimination than normal learning; that is, learning with novelty seeking. It is of interest to note that good discrimination capacity (such as perfect pitch) is often seen in individuals with autism (see, e.g., Frith, 1989).

Although our results show that restricted attention shifting may cause autistic-like characteristics of detailed learning in a narrow field, it is only the familiarity-preference hypothesis for restricted attention that is supported; the hypothesis of an attention shift impairment is not. We do not, however, claim to have refuted the hypothesis by Courchesne *et al.* that autism may be caused by attention-shift impairment. Their hypothesis builds on arguments involving multi-

modal sequences of stimuli—complex situations that have not been simulated here.

It should finally be stressed that the conclusions of this article—primarily that learning with familiarity preference often results in inadequate maps—do not diminish the possibility that inadequate maps may have such causes as those hypothesized in the works by Cohen, Gustafsson, McClelland, and Oliver *et al.*

ACKNOWLEDGMENTS

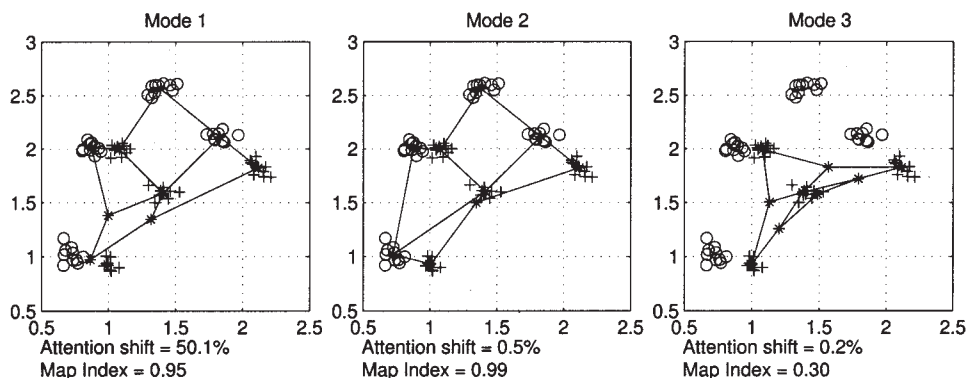**Fig. 8.** Learning data set 3 in a 3 × 3 SOM.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders.* (4th ed.). Washington, DC: Author.

Asperger, H. (1944). Die "autistischen Psychopathen" im Kindesalter. *Arch. Psychiatrie Nervenkrankheiten, 117,* 76–136. (Translated in Frith U. (Ed.) (1991). *Autism and Asperger Syndrome.* Cambridge: Cambridge University Press.)

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

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

Courchesne, E. (2002). Deciphering the puzzle: Unusual patterns of brain development in autism. In *Inaugural World Autism Congress.* Melbourne, Australia.

Courchesne, E., Akshoomoff, N., Townsend, J., & Saitoh, O. (1995). A model system for the study of attention and the cerebellum: Infantile autism. In G. Karmos, M. Molnár, I. Csépe, & J. Desmedt (Eds.), *Perspectives of event-related potentials research* (pp. 315–325). Amsterdam: Elsevier Science.

Courchesne, E., Townsend, J., Akshoomoff, N., Saitoh, O., Yeung-Courchesne, R., Lincoln, A., James, H., Haas, R., Schreibman, L., & Lau, L. (1994a). Impairment in shifting attention in autistic and cerebellar patients. *Behavioral Neuroscience, 108,* 848–865.

Courchesne, E., Townsend, J., Akshoomoff, N., Yeung-Courchesne, R., Press, G., Murakami, J., Lincoln, A., James, H., Saitoh, O., Egaas, B., Haas, R., & Schreibman, L. (1994b). A new finding: Impairment in shifting attention in autistic and cerebellar patients. In S. Broman & J. Grafinan (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain function* (pp. 101–137). Hillsdale, NJ: Erlbaum.

Dawson, G., Meltzoff, A., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders, 28,* 479–485.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford: Basil Blackwell.

Gillberg, C., & Coleman, M. (2000). *The biology of the autistic syndromes.* (3rd ed.). Cambridge University Press.

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

Happé, F. (1991). The autobiographical writings of three asperger syndrome adults: Problems of identification and implications for theory. In U. Frith (Ed.), *Autism and asperger syndrome* (pp. 207–242). Cambridge: Cambridge University Press.

Haykin, S. (1999). *Neural networks—A comprehensive foundation.* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hermelin, B. (1978). Images and language. In M. Rutter & E. Schoppler (Eds.), *Autism: A reappraisal of concept and treatment* (pp. 141–154). New York: Plenum.

Kandel, E., Schwartz, J., & Jessel, T. (Eds.). (2000). *Principles of neural science.* (4th ed.). New York: McGraw Hill.

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

Kohonen, T. (2001). *Self-organising maps.* (3rd ed.). Berlin: Springer.

Kootz, J., Marinelli, B., & Cohen, D. (1982). Modulation of response to environmental stimulation in autistic children. *Journal of Autism and Developmental Disorders, 12,* 185–193.

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

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

Oliver, A., Johnson, M. H., & Pennington, B. (2000). Deviation in the emergence of representations: A neuroconstructivist framework for analysing developmental disordere. *Developmental Science, 3,* 1–23.

Pascualvaca, D., Fantie, B., Papageorgiou, M., & Mirsky, A. (1998). Attentional capacities in children with autism: Is there a general deficit in shifting focus? *Journal of Autism and Developmental Disorders, 28,* 467–478.

Price, D. J., & Willshaw, D. J. (2000). *Mechanisms of cortical development.* Oxford: Oxford University Press.

Ritter, H., Martinetz, T., & Schulten, K. (1992). *Neural computation and self-organizing maps.* Reading, MA: Addison-Wesley.

Spitzer, M. (1995). A neurocomputational approach to delusions. *Compr. Psychiatry, 36,* 83–105.

Townsend, J., Harris, N., & Courchesne, E. (1996). Visual attention abnormalities in autism: Delayed orienting to location. *Journal of the International Neuropsychological Society, 2,* 541–550.

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