Encoding Shape and Spatial Relations: The Role of Receptive Field Size in Coordinating Complementary Representations

Robert A. Jacobs
Department of Psychology
University of Rochester
Rochester, NY 14627

Stephen M. Kosslyn
Department of Psychology
Harvard University
Cambridge, MA 02138

Date: November 1993
This paper has been accepted for publication in the journal *Cognitive Science*

*This research was supported by the McDonnell-Pew Program in Cognitive Neuroscience through JSMF Grant 90-33 and by NSF grant number IRI-9013991 to the first author, and by NSF Grant BNS 90-09619 and AFOSR Grant 91-0100 to the second author. Correspondence can be sent to either author.

Abstract

An effective functional architecture facilitates interactions among subsystems that are often used together. Computer simulations showed that differences in receptive field sizes can promote such organization. When input was filtered through relatively small nonoverlapping receptive fields, artificial neural networks learned to categorize shapes relatively quickly; in contrast, when input was filtered through relatively large overlapping receptive fields, networks learned to encode specific shape exemplars or metric spatial relations relatively quickly. Moreover, when the receptive field sizes were allowed to adapt during learning, networks developed smaller receptive fields when they were trained to categorize shapes or spatial relations, and developed larger receptive fields when they were trained to encode specific exemplars or metric distances. In addition, when pairs of networks were constrained to use input from the same type of receptive fields, networks learned a task faster when they were paired with networks that were trained to perform a compatible type of task. Finally, using a novel modular architecture, networks were not pre-assigned a task, but rather competed to perform the different tasks. Networks with small nonoverlapping receptive fields tended to win the competition for categorical tasks whereas networks with large overlapping receptive fields tended to win the competition for exemplar/metric tasks.

Encoding Shape and Spatial Relations: The Role of Receptive Field Size in Coordinating Complementary Representations

The brain is not a single, undifferentiated neural network. Rather, it clearly has a modular structure, and the individual subsystems interact in complex ways (e.g., see Desimone and Ungerleider, 1989; Felleman and Van Essen, 1991). Many Artificial Intelligence researchers have pointed out the virtues of modular architectures (e.g., Marr, 1982; Simon, 1981), and the modular organization of the brain may reflect the operation of fundamental computational principles. Specifically, distinct subsystems may have developed to carry out incompatible input/output mappings (see Kosslyn and Koenig, 1992). However, subsystems often are not entirely independent; rather, many may work together to accomplish most tasks. If so, then an effective computational architecture will "yoke" subsystems that often operate together, facilitating their joint operation. In this article we propose a simple mechanism that will serve to yoke at least some subsystems used in visual perception.

Visual perception relies on two major systems, which are localized in different parts of the brain (e.g., see Mishkin, Ungerleider, and Macko, 1983). The "dorsal system" runs up from the occipital lobe to the parietal lobe, and encodes spatial properties such as location, orientation, and size. In contrast, the "ventral system" runs from the occipital lobe down to the inferior temporal lobe, and this system encodes object properties such as shape, color and texture (for a review, see Chapter 3 of Kosslyn and Koenig, 1992). One virtue of this architecture is that it allows the ventral system to ignore location in the field during object recognition while at the same time the dorsal system preserves this information for other purposes (cf. Gross and Mishkin, 1977).

These two large systems can be broken down into sets of more specialized subsystems. First, let us consider two more fine-grained subsystems in the dorsal system. Conceptually, there is a clear distinction between *categorical spatial relations*, such as above/below, left/right, and on/off, and *coordinate spatial relations* that specify locations in a way that can be used to guide precise movements. Categorical spatial relations group a range of positions and treat them as equivalent; such representations are an essential feature of structural descriptions, which specify the arrangement of an object's parts in a way that applies to all of the object's various shape configurations. In contrast, metric coordinate spatial relations specify the information that is discarded by categorical relations. Such precise spatial information is essential for reaching and navigation. This is a good example of incompatible input/output mappings: To encode categorical spatial relations, a subsystem must discard the very information that is required to encode coordinate spatial relations.

This conceptual distinction suggests that distinct subsystems may encode the two types of spatial relations. One way in which researchers have established the functional distinction between two subsystems is rooted in the logic of a "double dissociation" (Teuber, 1955). In this case, the aim is to demonstrate that one subsystem operates more effectively in one cerebral hemisphere whereas the other operates more effectively in the other cerebral hemisphere. If there were only one way to encode spatial relations, either one of the hemispheres would be generally better or there would be no difference between the hemispheres; hence, if one hemisphere is better at encoding one sort of spatial relation, but the other hemisphere is better at another sort, this is good evidence for the existence of distinct subsystems. Many experiments have now demonstrated that the left cerebral hemisphere encodes categorical spatial relations (above/below, right/left, and on/off) more effectively than the right hemisphere (although this effect is often small in a given experiment), but the right cerebral hemisphere encodes metric coordinate spatial relations more effectively than the left hemisphere (for a review and meta-analysis, see Kosslyn, Chabris, Marsolek, and Koenig, 1992).

Similarly, there is evidence that the ventral system also can be decomposed into at least two encoding subsystems. Conceptually, there is a clear distinction between recognizing a stimulus as a member of a category (e.g., a dog) versus recognizing it as a specific exemplar (Fido). A category groups various exemplars and treats them as equivalent, whereas identifying an exemplar requires treating the instances as distinct. Again, the two mappings are incompatible: The very information that is needed to specify a specific example must be ignored to assign it to a category. And again, there is evidence that the cerebral hemispheres are specialized for the different types of encoding. For example, Marsolek, Kosslyn, and Squire (1992) asked subjects to read a list of words and to rate the degree to which they liked each word. This was a cover task, intended only to lead the subjects to look at each word. Following this, the subjects saw word stems (e.g., "CAS___"), and were asked to complete the stems to form the first word that came to mind (e.g., "CASTLE"). The stems were presented in the left visual field (and hence were seen initially by the right cerebral hemisphere) or in the right visual field (and hence were seen initially by the left cerebral hemisphere). The words were divided into two lists, only half of which were shown to a given subject at the outset; each list was shown to half of the subjects. “Priming” was measured by observing how many words were completed to form the words on the corresponding list when it was seen initially, compared to the number of words that were completed to form words on the other list. More interesting, there was greater priming if the words seen initially were the same typographic case as the stem, and this advantage was totally confined to the right hemisphere.

This result indicates that the right hemisphere represents specific exemplars better than the left hemisphere. However, some priming still occurred in the left hemisphere, but equivalent amounts of priming occurred when the initial and test cases were in the same or in different typographic case. Thus, the left hemisphere could represent information in visual categories. Marsolek (1992) went on to show that the left hemisphere actually is better than the right when prototypes of dot patterns must be abstracted and matched to novel stimuli. Again, then, we find a double dissociation, which indicates that different processes match input to representations of exemplars or visual categories.

We were struck by the fact that categorical spatial relations and category representations of shape are encoded more effectively in the left hemisphere, whereas coordinate spatial relations and exemplar representations of shape are encoded more effectively in the right cerebral hemisphere. This arrangement makes sense; indeed, the purposes of the different types of spatial relations representations can only be accomplished effectively in the context of the appropriate types of representations of shape. An effective structural description will generalize to all instances of an object; this requires not only that the arrangement of parts be specified in a way that is robust over contortions of the object, but also that the parts themselves be represented in a way that will generalize over the variations in the shapes of parts. Similarly, to reach or navigate effectively, one not only needs to know how far away an object is, but also needs to know its precise shape; one will move around a table differently if it is square or circular.

It seems clear that an effective computational architecture will facilitate interaction between the complementary types of processing. But how does such coordination occur? Given the large amount of plasticity in the developing brain (e.g., see Dennis and Kohn, 1975; Dennis and Whitaker, 1976), the subsystems may not be innately configured in this way. Indeed, Kosslyn, Koenig, Brown, and Gazzaniga (cited in Chapter 9 of Kosslyn and Koenig, 1992) describe a split-brain patient in whom the usual pattern of laterality of spatial relations encoding, as inferred from divided-visual-field studies in normal subjects, appears to be reversed.

Kosslyn et al. (1992) showed that a simple property of neural information processing is capable of producing the observed left-hemisphere specialization for categorical spatial relations and the observed right-hemisphere specialization for metric coordinate spatial relations. In their neural network simulations, networks were trained to make a categorical judgment (whether a dot was above or below a bar) or a metric coordinate judgment (whether a dot was within four metric units of the bar). The networks could be trained to encode categorical spatial relations more effectively when the input units had relatively small, nonoverlapping receptive fields. These small receptive fields apparently helped the network to delineate pockets of space and to specify the spatial relations with respect to these pockets. In contrast, the networks could be trained to encode metric coordinate spatial relations more effectively when the input units had relatively larger, overlapping receptive fields. These large overlapping fields promoted the use of "coarse coding" to precisely encode the metric relation between a dot and bar.

There is much evidence that is consistent with the idea that the left hemisphere monitors the outputs of neurons with relatively small receptive fields whereas the right hemisphere monitors the outputs of neurons with relatively large receptive fields. Van Kleeck (1989) reports a meta-analysis of studies in which subjects are asked to find a target when they are shown large letters that are made by arranging copies of a small letter. They classify a smaller, component letter faster if the stimulus is presented initially to the left hemisphere, but classify the larger, overall letter faster if the stimulus is presented initially to the right hemisphere. Moreover, patients with damage to the left posterior superior temporal lobe have difficulty identifying component letters of hierarchical stimuli, whereas patients with damage to the right posterior superior temporal lobe have difficulty identifying the overall pattern formed by the smaller letters (e.g., Robertson and Lamb, 1991; Robertson, Lamb, and Knight, 1991). Similarly, subjects categorize high spatial frequency gratings faster if they are presented to the left hemisphere than to the right, but vice versa for low spatial frequency gratings (e.g., see Christman, Kitterle, and Hellige, 1991; Kitterle and Selig, 1991). Small receptive fields would detect finer variations better than large receptive fields, and vice versa for larger spatial variations.

The present research was designed to test the hypothesis that differences in the sizes of receptive fields can coordinate complementary representations of shape and spatial relations. We hypothesized that relatively small, nonoverlapping receptive fields promote the development of both categorical spatial relations representations and representations of shape categories, whereas relatively large, overlapping receptive fields promote the development of both metric spatial relations representations and representations of individual shapes.

General Method

We used four tasks in the experiments reported in this article. Two tasks required a system to make categorical judgments (these tasks are labeled *cat*) and two tasks required coordinate judgments (these tasks are labeled *coo*). Within these four tasks, two involved judging the spatial relation of one shape with respect to another (these tasks are labeled *where*) and two involved judging the identity of a shape (these tasks are labeled *what*). Specifically, a network specified whether a shape was above or below a horizontal bar when performing the where/cat task, and specified whether a shape was within two metric units of the bar or whether it was more than two units away when performing the where/coo task. The what/cat task required a network to specify the category of a shape, and the what/coo task required it to specify a shape's individual identity.

Each shape category was defined by a prototype, which was a pattern formed in a $5X5$ grid. By selecting one pixel of a category's prototype and perturbing that pixel one unit

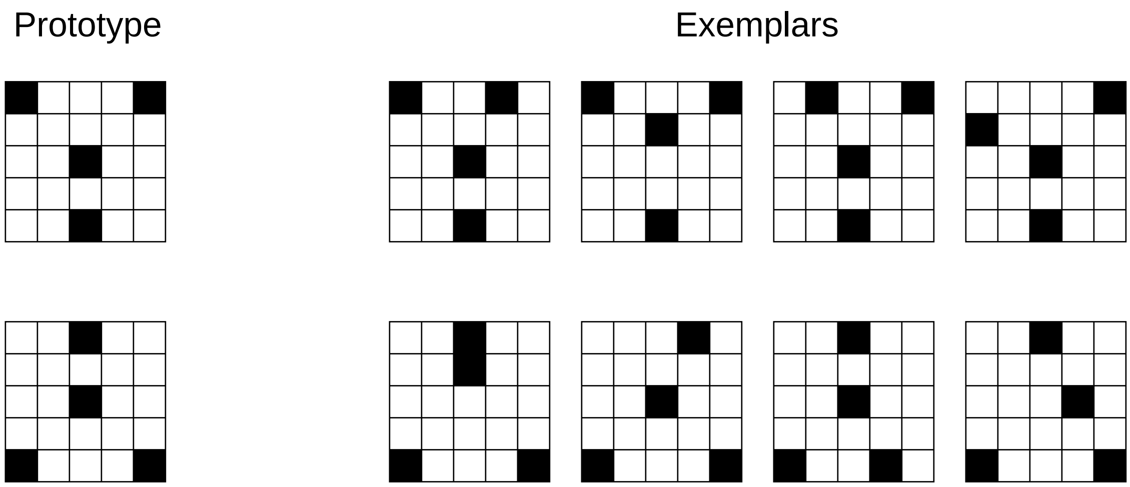

Figure 1: Prototypes and exemplars for the two shape categories.

left, right, up, or down, an exemplar from a category was formed. The experiments used two category prototypes and four exemplars from each category (see Figure 1).

The training of the networks was characterized by two time frames. One training pattern was presented to a network during each step; all training patterns were presented to a network exactly once during each epoch. The input array had 19 rows and 5 columns. The horizontal bar, which was represented by pixels in the first and fifth columns, was randomly placed at one of three rows in the center of the array at each step. One of the eight shapes was randomly selected and placed between one and four rows above or below the bar. The left side of Figure 2, for example, shows a shape located four units below the bar. Within each epoch of training, each of the eight shapes was presented at each of the eight locations with respect to the bar for all three possible positions of the bar. Thus there were 192 steps per epoch (8 shapes x 8 locations of the shapes x 3 locations of the bar).

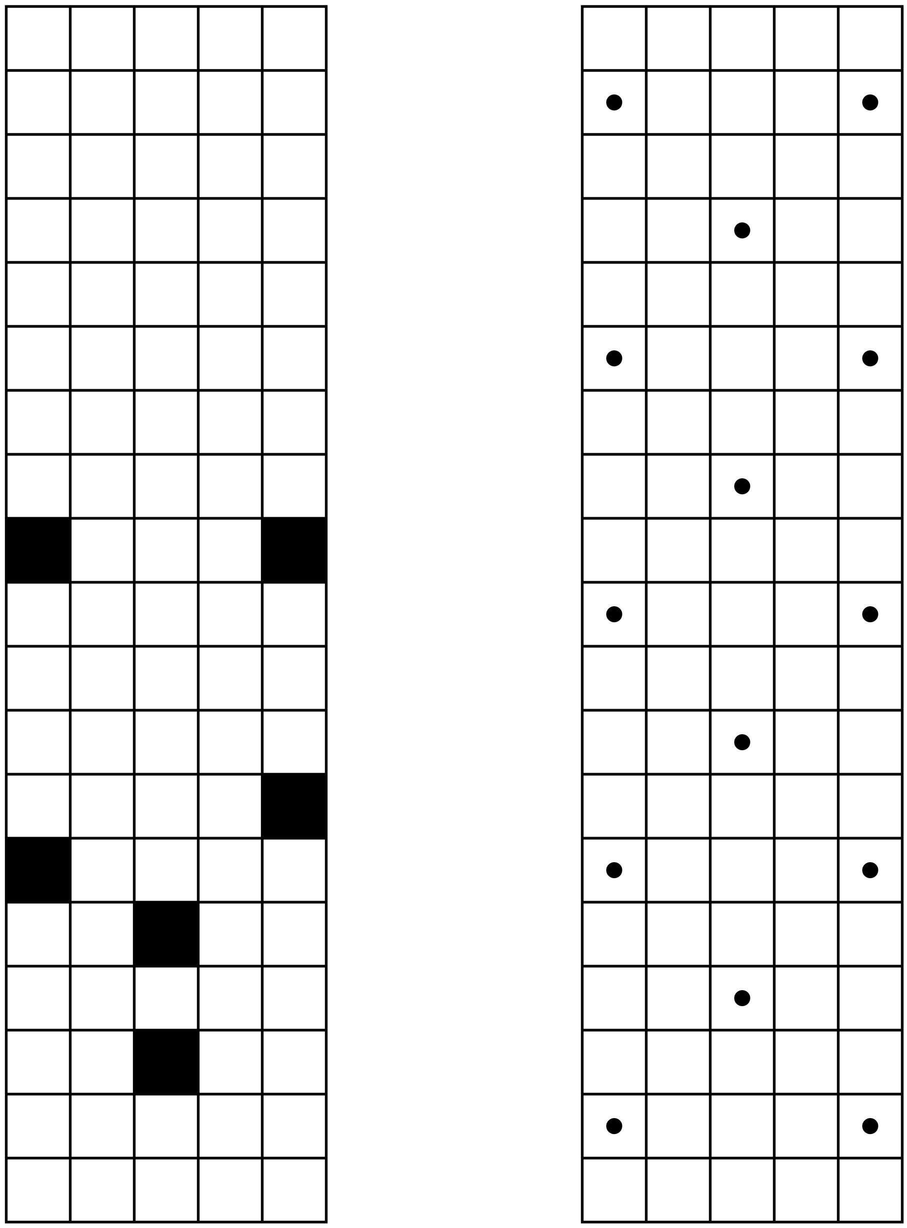

Figure 2: On the left is the input array with an exemplar located four units below the bar. On the right is the input array with dots showing the centers of the 14 Gaussian units.

The networks did not receive the array directly as input, but instead received a filtered version of the array. This filtering was performed by 14 Gaussian units. A unit's mean determined the center of its receptive field and its covariance matrix determined the receptive field's shape. As illustrated on the right side of Figure 2, the means of the Gaussian units were fixed so that the units formed a uniform tesselation over the input array (in this figure, units were centered at squares of the array containing a dot). All units of a network had identical covariance matrices which were fixed to be of the form $\sigma^2\mathbf{I}$ where σ^2 is a variance and $\mathbf{I}$ is the identity matrix. Thus all units had spherical receptive fields in which the variance determined the field's width.

The variance of the Gaussian units was set so that there was considerable overlap of the units' receptive fields. Consequently, the units provided the networks with a coarse-coded representation of the retinal array (Ballard, 1986; Hinton, 1981; Hinton, McClelland, and Rumelhart, 1986). An important feature of such codes is that their resolution depends on the degree of overlap. High resolution codes are formed by units with large, overlapping receptive fields whereas units with small, less-overlapping receptive fields form low resolution codes. A qualitative understanding of the relationship between receptive field size and resolution can be gained from Figure 3, which superimposes the input array and the Gaussian units' receptive fields. Whereas the units on the left have small receptive fields ($\sigma^2 = 0.8$) and provide a relatively low resolution code, the units on the right have large receptive fields ($\sigma^2 = 1.8$) and provide a high resolution code (for graphical purposes, the receptive field of a unit is shown as a circle whose radius is 2σ)[1]. Analyses by Ballard (1986) and Hinton (1981) provide a quantitative understanding of the relationship between receptive field size and resolution in the case of binary units that each become active when a stimulus falls within its receptive field. If D is the diameter of a unit's receptive field, k is the dimensionality of the space to be represented, and N is the desired number of just-noticeable differences in each dimension (i.e. the desired resolution), then the required number of units is N^k/D^{k-1}. Note that for a fixed number of units, a high resolution code (that is, one with a large N) requires units with large receptive fields (that is, fields with a large D) whereas a low resolution code can be achieved with units with small receptive fields.

The total activation of a Gaussian unit was determined by first computing its partial activations due to the presence of individual pixels in the retinal array. Define the partial activation of a Gaussian unit as

$$x_i = \frac{1}{\sigma^2} e^{-\frac{1}{2\sigma^2}||p_i - \mu||^2} \tag{1}$$

where p_i is the location of the i^{th} pixel that is on in the retinal array and μ is the mean of the Gaussian unit. The total activation of a unit was the sum of its partial activations; that is

$$x = \sum_i x_i$$

[1] Strictly speaking, Gaussian units (with $\sigma^2 > 0$) have infinite support and, thus, always have fully overlapping receptive fields. In addition, the activations of Gaussian units are real-valued and therefore provide infinite resolution. Nevertheless, the difficulty of distinguishing between two nearby real-values means that the arguments presented here are correct in practice, though not in theory.

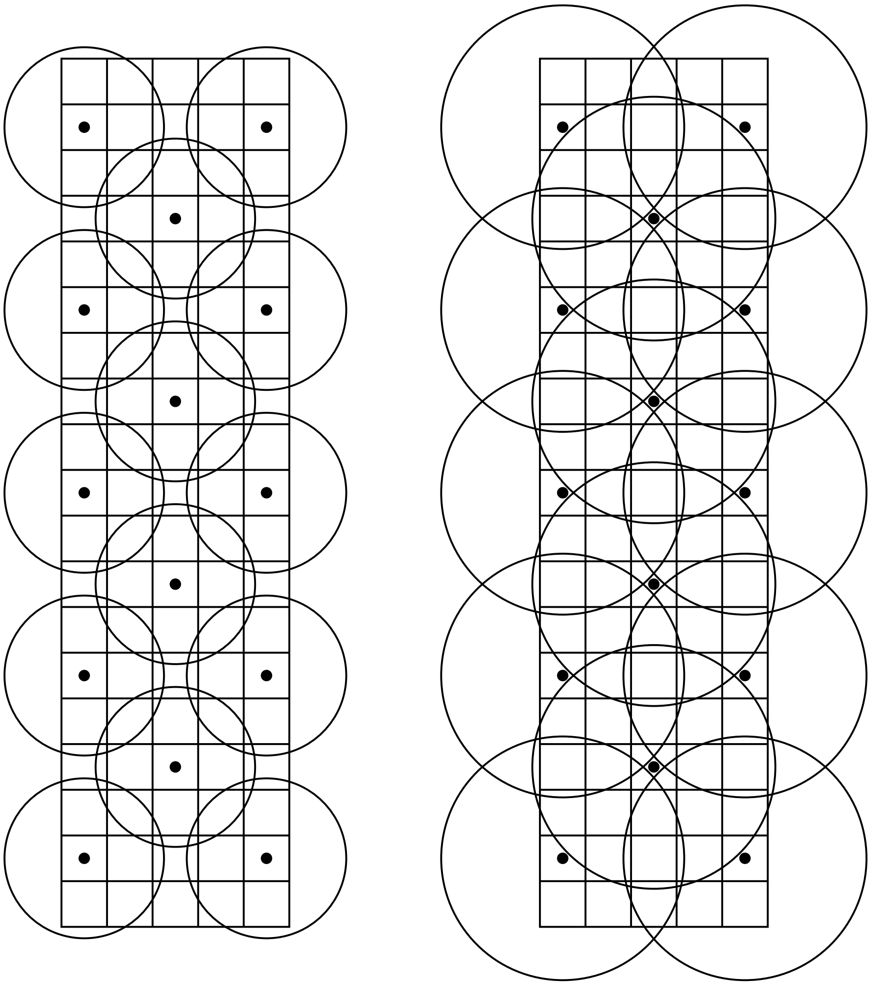

Figure 3: The Gaussian units' receptive fields superimposed on top of the input array. The small receptive fields ($\sigma^2 = 0.8$) of the units on the left provide a low resolution representation of the array whereas the large receptive fields ($\sigma^2 = 1.8$) of the units on the right provide a high resolution representation. For graphical purposes, the receptive field of a unit is shown as a circle whose radius is 2σ.

where i ranges over the number of pixels that were on in the array.

The networks contained three layers of units which were connected in a strictly feedforward manner. The 14 Gaussian units formed the input layer. The hidden and output layers contained units that used the logistic activation function. The hidden layer had 20 units, and the number of units in the output layer depended upon the task; it contained 2 units for the where/cat, where/coo, and what/cat tasks, and contained 8 units for the what/coo task. The weights of the networks were adjusted during training so as to minimize the sum of square error cost function

$$E = \sum_p \sum_i (y^*_{pi} - y_{pi})^2 \tag{2}$$

where y^*_{pi} is the target activation for the i^{th} output unit on the p^{th} training pattern and y_{pi} is the i^{th} output unit's actual activation on that pattern. The weight adjustment was determined by a conjugate gradient optimization procedure (Press, Flannery, Teukolsky, and Vetterling, 1986) in which the derivatives of the cost function E with respect to the weights were computed using the backpropagation algorithm (Rumelhart, Hinton, and Williams, 1986).

Experiment 1

In the first experiment, we tested the most basic prediction of our hypothesis, namely that there should be an interaction between receptive field size and the speed of learning the different tasks. Specifically, the categorical tasks should be learned faster with small, nonoverlapping receptive fields whereas the coordinate/exemplar tasks should be learned faster with large, overlapping receptive fields. Three networks were trained to perform each of the four tasks. The first network had Gaussian units with relatively small receptive fields ($\sigma^2 = 0.8$), the second network had moderate size receptive fields ($\sigma^2 = 1.3$), and the third network had large receptive fields ($\sigma^2 = 1.8$).

The results are shown in the four graphs of Figure 4. The graphs' vertical axes give the average number of epochs required for a network to successfully learn to perform a task (the error bars indicate the standard deviation). Our criterion for success was that the network correctly performed 95% of a task's training patterns. A training pattern was performed correctly when each output unit's activation was greater than 0.6 whenever its target value was 1 and was less than 0.4 whenever its target value was 0. The graphs' horizontal axes give the receptive field size. Twenty runs of each network on each task were performed. The number of runs that converged to a correct solution is also provided along the horizontal axes.

For the where/cat task (the network must decide whether a shape is above or below the horizontal bar), networks learned equally rapidly despite the differences in receptive field size. This task was learned so rapidly that it did not permit the advantages and disadvantages of different size receptive fields to manifest themselves. Fortunately, this was not the case for the other three tasks. For the where/coo task (the network must decide if a shape is more or less than two units away from the bar), networks with moderate size or large receptive fields learned significantly faster than networks with small receptive fields ($t = 10.30$, $p <$

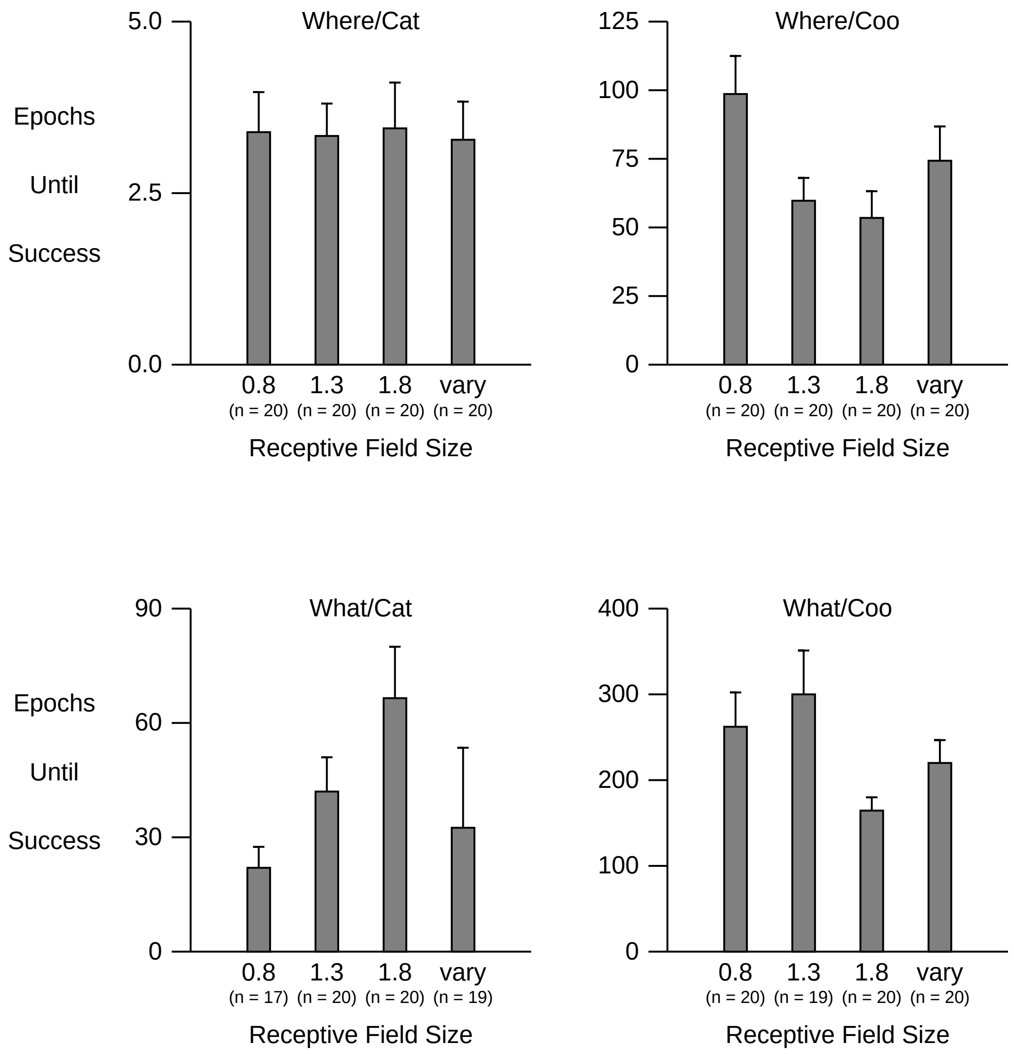

Figure 4: Learning speeds for networks with different size receptive fields on the four tasks. The vertical axes give the average number of epochs required for a network to successfully learn to perform a task (the error bars indicate the standard deviation). The graphs' horizontal axes give the receptive field size. Twenty runs of each network on each task were performed. The number of runs (n) that converged to correct solutions is also provided along the horizontal axes. The rightmost bar of each graph gives the learning speeds for networks whose receptive field sizes could adapt during training (see Experiment 2).

0.01 and t = 11.41, $p < 0.01$ respectively). Kosslyn et al. (1992) reported similar results. For the what/cat task (the network must specify the category of a shape), networks with small receptive fields learned faster than those with moderate size receptive fields (t = 7.68, $p < 0.01$), which in turn learned faster than those with large receptive fields (t = 6.56, p < 0.01). In contrast, for the what/coo task (the network must specify a shape's individual identity), networks with the largest receptive fields learned faster than those with moderate or small size receptive fields (t = 11.28, $p < 0.01$ and t = 9.85, $p < 0.01$ respectively). The results are, on the whole, consistent with our predictions. Whereas shape categorization was learned faster by networks with small receptive fields, coordinate/exemplar tasks were learned faster by networks with large receptive fields.

Experiment 2

In this experiment we tested networks in which the receptive field sizes as well as the weights could adapt. Our prediction was that the receptive fields of networks trained to perform categorical tasks would become small whereas the receptive fields of networks trained to perform coordinate tasks would become large. The process for adapting receptive field sizes is identical to the process for adapting weights. The derivative of the cost function E (Equation 2) with respect to the variance σ^2 of the Gaussian units was computed using the backpropagation algorithm. This derivative was then used by a conjugate gradient optimization procedure to adapt the value of the variance. Networks were initialized with moderate size receptive fields ($\sigma^2 = 1.3$).

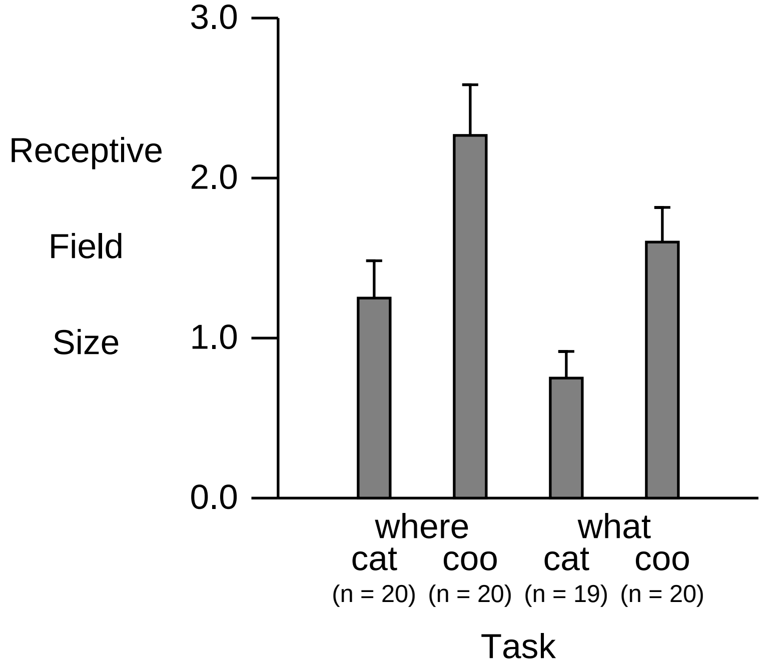

Figure 5: Receptive field sizes (mean and standard deviation) of networks after learning to perform the four tasks. The number of runs (n out of 20) that converged to correct solutions is provided along the horizontal axis.

The networks' receptive field sizes for the four tasks at the point when they had successfully learned to perform each task are shown in Figure 5. Consistent with our earlier findings, receptive fields of networks trained to perform the what/cat task became relatively small whereas receptive fields of networks trained to perform the where/coo and what/coo tasks became large (all pairwise differences among the four groups are statistically significant at the $p < 0.01$ level). The where/cat task again was so easy that it was rapidly learned using the moderate size receptive field with which networks were initialized.

The learning speeds for these networks are given by the rightmost bars in the four graphs in Figure 4, which allows one to compare the learning speeds of the networks in Experiments 1 and 2. The results suggest that, for a given task, networks with adaptive receptive fields learn faster than networks whose receptive fields are inappropriately sized for the task, but slower than networks with appropriately sized receptive fields. The adaptive receptive field networks generally learned the categorical tasks faster than networks with fixed large receptive fields, although not as fast as networks with small receptive fields. For coordinate tasks, the opposite result was found. These networks learned the coordinate tasks faster than networks with small receptive fields but learned slower than networks whose receptive fields were large.

As a strategy for nervous systems, these results suggest that the ability to adapt receptive field sizes is conservative. It has the advantage of allowing networks to tailor their structure to the nature of the task that they need to perform, and thereby achieve good performance. However it has the disadvantage that networks do not achieve the optimal performance that would occur if biology had hardwired them with suitable structures.

We reported above the receptive field sizes of networks at the point when the networks had successfully learned to perform each of the four tasks. We have also examined how the receptive field sizes changed during the course of learning. In this case, networks trained on the categorical tasks were initialized with small receptive fields ($\sigma^2 = 0.8$) and networks trained on the coordinate tasks were initialized with large receptive fields ($\sigma^2 = 1.8$).

Figure 6 shows the results for the two categorical tasks. The vertical axis provides the receptive field size averaged over ten networks. The horizontal axis provides the training time in epochs. For both the where/cat and what/cat tasks, the receptive field sizes remained stable at a small value.

The results for the two coordinate tasks are shown in Figure 7. For the where/coo task, the receptive field size grew larger than the relatively large value with which it was initialized. The receptive field size of networks trained on the what/coo task, in contrast, grows small during the early stages of training but then returns to a relatively large value. This suggests that acquisition of the what/coo task is facilitated by the use of a succession of representations. Early in training, relatively low resolution representations allow a network to learn the coarse features of the shapes. As training progresses, higher resolution representations permit a network to learn finer details. This learning strategy is reminiscent of the use of multi-resolution pyramids in many computer vision systems (Ballard and Brown, 1982).

Experiment 3

If biological neural networks in the same cerebral hemisphere share a common receptive field size, then there ought to be computational advantages to placing networks that perform the same type of task within the same hemisphere. That is, it should be advantageous to group the networks that perform categorical tasks in one hemisphere, and provide them with a relatively low resolution representation of the visual image. Similarly, it should be advantageous

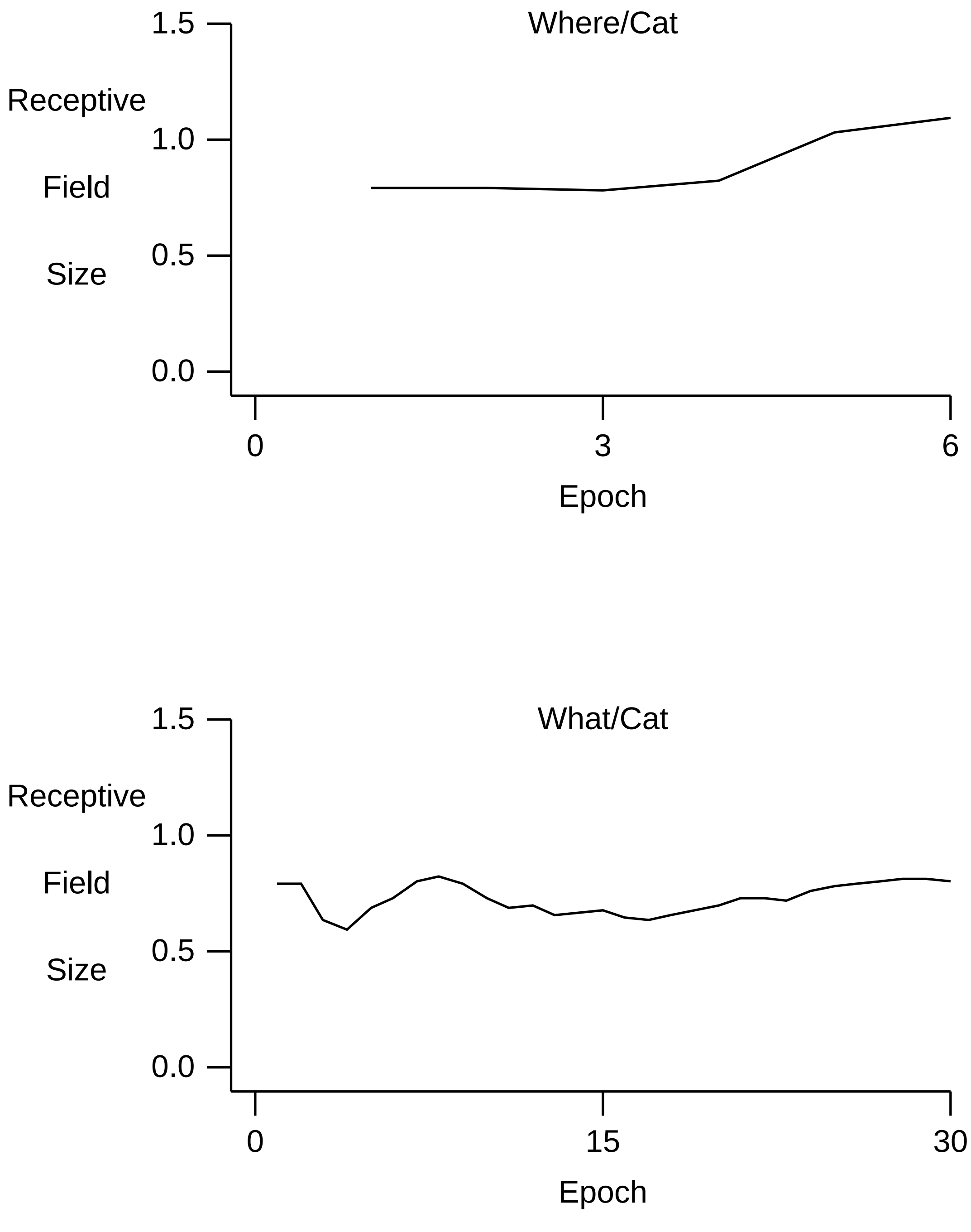

Figure 6: The average receptive field sizes at each epoch of training on the categorical tasks.

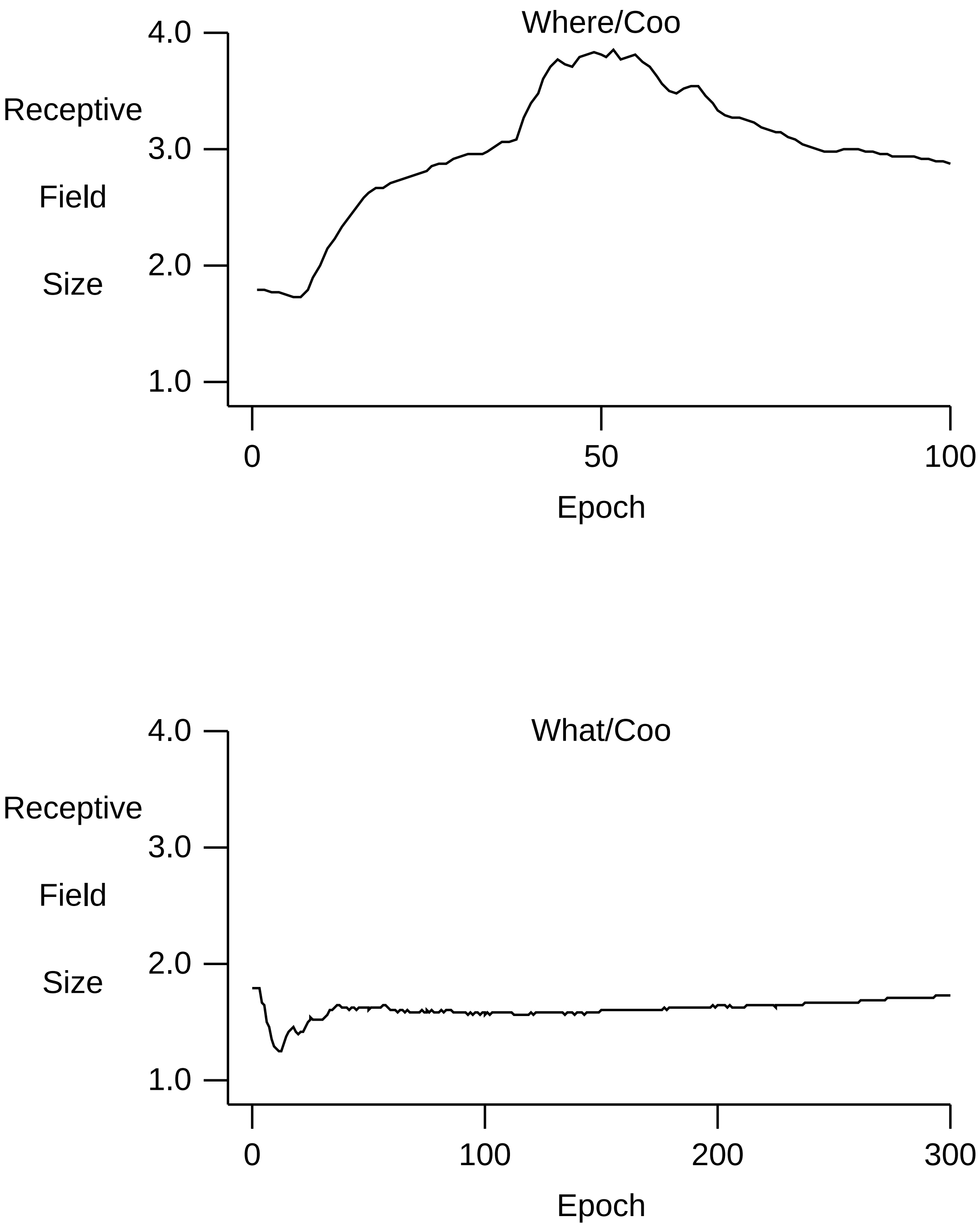

Figure 7: The average receptive field sizes at each epoch of training on the coordinate/exemplar tasks.

to group the networks that perform coordinate/exemplar tasks in the other hemisphere, and provide them with a relatively high resolution representation[2]. We investigated this possibility in Experiment 3.

The systems that we studied consisted of pairs of networks that were trained to perform pairs of tasks. The two networks shared a common input layer; that is, they received the activations of the same Gaussian units as inputs. Otherwise they were independent. During training, both networks sent error information to the Gaussian units. These units used this information to adapt their receptive field sizes.

The two-network systems were trained on four pairs of tasks. The pairs all consisted of a "where" task and a "what" task, although they differed in their combination of categorical and coordinate tasks. As is shown in Table 1, a network learned a task faster when it was paired with a network that was trained to perform a similar type of task. For example, a network trained to perform the where/coo task required less time to learn when it was paired with a network trained on the what/coo task than when it was paired with a network trained on the what/cat task ($t = 2.64$, $p < 0.01$). As another example, the learning speed of a network trained to perform the what/cat task was greater when it was paired with a network trained on the where/cat task than when it was paired with a network trained on the where/coo task ($t = 7.50$, $p < 0.01$). Consistent with the arguments given above, there

[2] An alternative, which does not appear to be used by biological systems, is to group the networks performing "where" tasks in one hemisphere and the networks performing "what" tasks in the other hemisphere.

Table 1: The learning speeds (mean and standard deviation) for the two-network systems. The number of runs (n out of 20) that converged to correct solutions is provided in the rightmost column.

Tasks	Epochs Until Success		n
	Mean	St. Dev.	
where/cat	3.60	0.75	20
what/cat	34.50	6.82	
where/cat	4.89	0.81	18
what/coo	322.55	44.62	
where/coo	108.15	19.35	20
what/cat	61.70	14.71	
where/coo	92.35	18.55	20
what/coo	230.25	21.65	

are indeed computational advantages to grouping networks that perform similar types of tasks.

For each of the four pairs of tasks, Figure 8 provides the receptive field sizes at the point when both tasks were successfully learned. With two categorical tasks, not surprisingly, the receptive fields became small whereas with two coordinate tasks, the receptive fields grew large. With two of the dissimilar tasks, the where/coo and what/cat tasks, the receptive fields assumed a moderate size. However with the where/cat and what/coo tasks, the receptive fields became large. This is probably due to the fact that the where/cat task can be rapidly learned with a wide variety of receptive field sizes (see Figure 4).

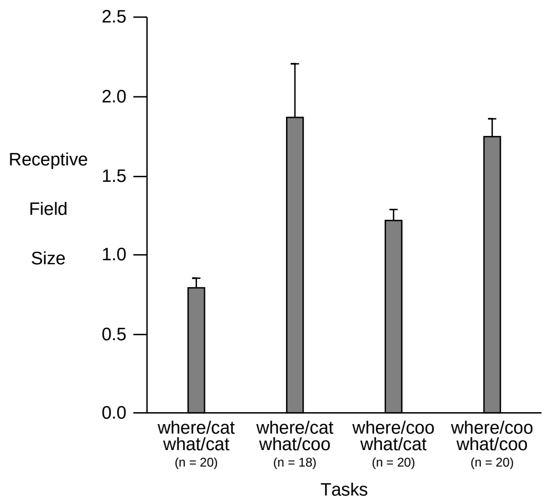

Figure 8: Receptive field sizes (mean and standard deviation) of two-network systems after learning the four pairs of tasks. The number of runs (n out of 20) that converged to correct solutions is provided along the horizontal axis.

Experiment 4

In the experiments reported above we showed that the nature of a task can influence a network's structure, in this case, its receptive field size. In Experiment 4 we looked for the converse affect. That is, can a network's structure influence the nature of the task that the network learns to perform? To answer this question, we used a multi-network, or modular, architecture in which tasks are not pre-assigned to networks. Instead networks compete for the "right" to learn different tasks. Previous research has shown that the structure of each network biases the competition such that networks tend to learn tasks for which their structure is well-suited (Jacobs, Jordan, and Barto, 1991). Based on our earlier findings, we predicted that networks with small receptive fields would tend to win the competition for categorical tasks whereas networks with large receptive fields would tend to win the competition for coordinate tasks.

The architecture we used was first described in Jacobs, Jordan, Nowlan, and Hinton (1991) and combines earlier work on learning in a modular architecture by Jacobs, Jordan, and Barto (1991) with the mixture models view of competitive learning advocated by Nowlan (1990) and Hinton and Nowlan (1990). As illustrated in Figure 9, it consists of two types of networks called *expert networks* and a *gating network*. The expert networks compete to learn each of the training patterns, and the gating network mediates this competition. Whereas the expert networks have an arbitrary structure, the gating network is restricted to have as many output units as there are expert networks, and the activations of these units must be nonnegative and sum to one. To meet these constraints, the output units use the "softmax" activation function (Bridle, 1989); specifically the activation of the i^{th} output unit of the gating network, denoted g_i, is

$$g_i = \frac{e^{s_i}}{\sum_{j=1}^{n} e^{s_j}} \tag{3}$$

where s_i denotes the weighted sum of unit i's inputs and n denotes the number of expert networks. The output of the entire architecture, denoted $\mathbf{y}$, is

$$\mathbf{y} = \sum_{i=1}^{n} g_i \mathbf{y}_i \tag{4}$$

where $\mathbf{y}_i$ denotes the output of the i^{th} expert network.

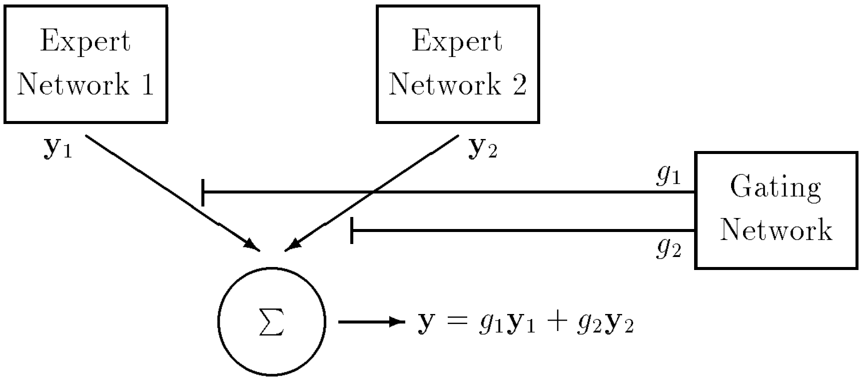

Figure 9: A modular connectionist architecture consisting of two expert networks and a gating network.

The learning process of the modular architecture combines aspects of competitive and associative learning. Consider the following intuitive description of the architecture's learning process (a more precise, mathematical description can be found in the Appendix). During training, the weights of the expert and gating networks are adjusted simultaneously. The output of each expert network is compared with the desired output at each time step. The expert whose output most closely matches the desired output is called the winner of the competition. The other experts are called losers. Each expert receives an amount of training information that is proportional to its relative performance on the training pattern. Whereas the winning expert network receives a lot of information, and thus learns a lot about the current training pattern, the losing expert networks receive little or no information, and thus learn little about the current pattern. The gating network adjusts its weights so as to increase the activation of the output unit corresponding to the winning expert and to decrease the activations of its remaining output units.

The learning process has a positive feedback effect that forces different expert networks to learn different tasks. This effect relies on the fact that, in general, training patterns from the same task share a common underlying structure whereas patterns from different tasks have different underlying structures. Suppose that at some time step, an expert has won the competition to learn some of the training patterns from one particular task. In this case, the expert will have at least partial "knowledge" of the structure of the task. In the future, therefore, it will be likely to win the competition for the remaining patterns from that task. The expert will thereby become specialized for performing the task. As a result of this specialization, however, this expert will be quite likely to perform poorly on patterns from other tasks—unless some tasks happen to be very similar. Thus other experts will be likely to win the competition for the patterns from other tasks. In this way, different experts win the competition to learn patterns from different tasks, and the experts become specialized for performing different tasks.

In the simulations we performed, the modular architecture had two expert networks. The structure of each expert was identical to the structure of the networks used in the previous experiments. Experts had an input layer comprised of the 14 Gaussian units, a hidden layer of 20 units, and an output layer of either 2 or 8 units. The gating network had 2 input units and 2 output units. The input units indicated which of two tasks the architecture should perform at each time step. The two tasks were either the where/cat and where/coo tasks, or the what/cat and what/coo tasks. The arrangement of inputs provided to the expert and gating networks forced the architecture to dedicate one expert to each task. However, it did not restrict which expert could learn each task[3].

The architecture's two expert networks differed in their receptive field sizes. For the where/cat and where/coo tasks, we tested two conditions. In the first condition, we gave one expert a small receptive field size ($\sigma^2 = 0.8$) and gave the other expert a large receptive field size ($\sigma^2 = 1.8$). In the second condition, we exaggerated even further the difference in receptive field sizes. One expert's receptive fields were very small ($\sigma^2 = 0.5$) whereas the other expert's were very large ($\sigma^2 = 2.1$).

Recall that we predicted that the architecture would discover an appropriate match between the structure of an expert and the structure of a task. In the first condition, on only 8 out of 20 runs did the expert with small receptive fields win the competition for the where/cat task and the expert with large receptive fields win the competition for the where/coo task. We attribute this disappointing result to the fact that the where/cat task can be learned rapidly by networks with a wide variety of receptive field sizes. In the second condition, where the difference in receptive field sizes between the expert networks was exaggerated, 18 out of 20 runs showed the predicted results (thus we can reject the null hypothesis that the assignment of networks to tasks is not biased in the predicted way, p <

[3]Note that the what/cat and what/coo tasks have desired output vectors of two and eight elements respectively. When trained on these tasks, both experts were given eight output units, and the desired output vector contained six "don't care" conditions on time steps when the what/cat task was to be performed (Jordan, 1986).

0.01). On these runs, the expert with very small receptive fields won the competition for the where/cat task whereas the expert with very large receptive fields won the competition for the where/coo task.

For the what/cat and what/coo tasks, we only examined the case where one expert's receptive fields were small ($\sigma^2 = 0.8$) and the other expert's receptive fields were large ($\sigma^2 = 1.8$). On 16 out of 20 runs the predicted outcome was achieved (again we can reject the null hypothesis that the assignment of networks to tasks is not biased in the predicted way, $p < 0.01$). These results show that the architecture is sensitive to discovering appropriate matches between the structure of an expert and the structure of a task.

Analysis of Coarse-Coded Representations

The results of the four experiments strongly suggest that Gaussian units with different receptive field sizes possess different functional properties. In order to better understand these results, we have analyzed the representations formed by these units in a more direct manner. Above we characterized the codes formed by Gaussian units with different size receptive fields as possessing either low or high resolution. In doing so, we are describing how these codes differ in the level of detail that they provide about individual patterns on the retinal array. It is also possible to characterize the differences between these codes in terms of the similarity or dissimilarity of the representations that they assign to different patterns on the array. For example, if a network is learning to perform the what/cat task (specify the category of a shape), then it is likely to be useful if different exemplars from the same shape category are assigned similar representations by the Gaussian units, and if exemplars from different categories are assigned dissimilar representations. Alternatively, if a network is learning to perform the what/coo task (specify a shape's individual identity), then it is likely to be useful if the Gaussian units assign dissimilar representations to all the exemplars.

In response to the placement of a visual pattern on the retinal array, each of the fourteen Gaussian units computes an activation level. That is, each visual pattern is represented by a point in a fourteen-dimensional space, which we call GU space (Gaussian unit space). Note that Gaussian units with different receptive field sizes represent the same set of visual patterns using different points. We have measured the Euclidean distances between these points when the receptive field sizes are either small or large.

Define an output set to be the collection of visual patterns that require identical responses for a given task. For example, there are two output sets for the where/cat task: one set consists of the patterns in which the shape is above the horizontal bar and the other set consists of the patterns in which the shape is below the bar. As a useful heuristic, define a measure of the "goodness" of a code to be the minimum Euclidean distance between points in GU space corresponding to patterns from different output sets. If this distance is large, then a code is said to be good because different output sets in this code are far apart.

The top graph in Figure 10 gives the level of goodness for codes formed either by Gaussian units with small receptive fields ($\sigma^2 = 0.8$) or by units with large receptive fields ($\sigma^2 = 1.8$) for the four tasks. The horizontal axis gives the receptive field size and the task. The vertical axis gives the minimum distance between points corresponding to patterns from different output sets. The prototypes for the two shape categories are shown in the upper right corner. This graph illustrates several properties of codes formed by Gaussian units. Of primary importance for our purposes is that units with small receptive fields ($\sigma^2 = 0.8$) provide a better code for the categorical tasks whereas units with large receptive fields ($\sigma^2 = 1.8$) provide a better code for the coordinate tasks. Based on the results of the what/cat and what/coo tasks, this appears to be due to the fact that units with small receptive fields give highly dissimilar representations to exemplars from different categories but relatively similar representations to exemplars from the same category. Units with large receptive fields, on the other hand, give moderately dissimilar representations to all exemplars.

We wondered whether these results reflect the properties of Gaussian units with different receptive field sizes or are an idiosyncratic artifact of the particular visual stimuli that we have used. We created, therefore, a new set of visual stimuli in the exact manner as described above with the exception that there were four shape categories instead of two. The results are given in the bottom graph of Figure 10. The prototypes of the four shape categories are shown in the upper right corner of this graph. In short, the pattern of results is unchanged. Gaussian units with small receptive fields provide a better code for the categorical tasks whereas units with large receptive fields provide a better code for the coordinate tasks.

Our analysis of coarse-coded representations requires some clarifications. We are attempting to evaluate the hypothesis that Gaussian units with relatively small receptive fields are more appropriate for categorical tasks whereas units with relatively large receptive fields

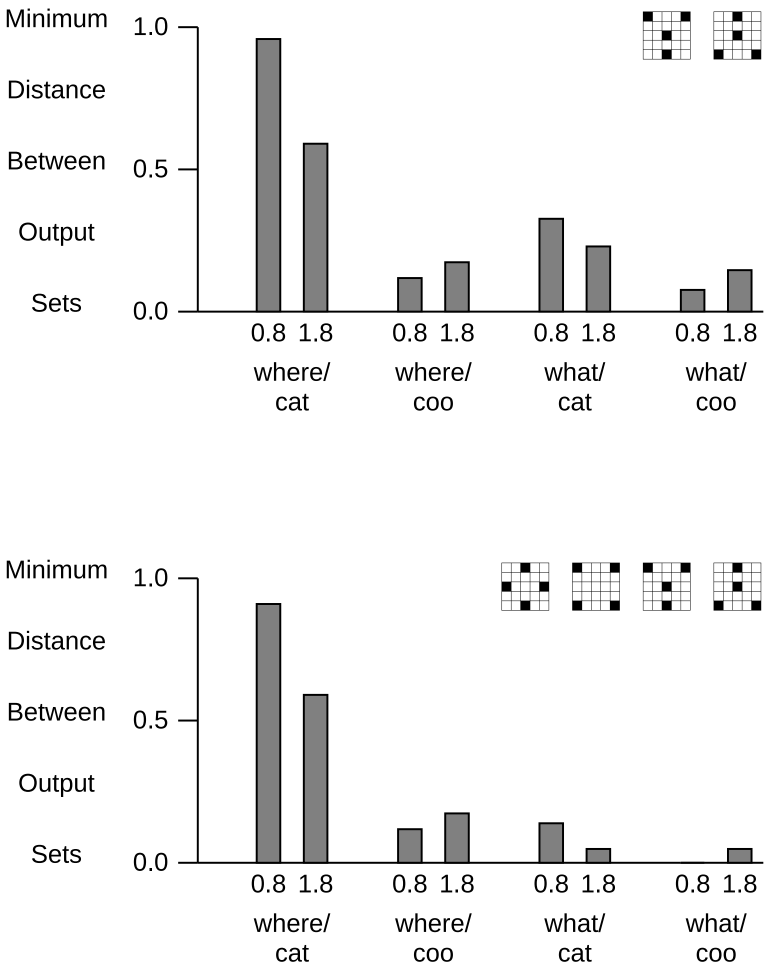

Figure 10: The vertical axes of these graphs give the minimum distance between Gaussian unit representations for patterns from different output sets. The horizontal axes give the size of the units' receptive fields and also the task. The top graph is for the case where the visual patterns were formed using two shape categories (these were the patterns that were used in all experiments). The bottom graph is for the case where the visual patterns were formed using four shape categories. The prototypes of the shape categories are shown in the upper right corner of each graph.

are more appropriate for coordinate tasks. The above analysis provides supporting evidence when the Gaussian units have receptive field sizes within a given range (variances of either 0.8 or 1.8), and when the visual stimuli are formed from shape categories whose prototypes have little or moderate overlap. One may, however, question the generality of these results. For example, the results might be different if the prototypes of the shape categories are highly similar. In this case, we expect that a comparison of units whose variance equals 0.8 with those whose variance equals 1.8 would show that the latter units provide a better code for both categorical and coordinate tasks. This does not mean, however, that our hypothesis is incorrect. The hypothesis concerns the *relative* receptive field sizes, and it predicts the existence of a range of receptive field sizes with the desired properties. For example, it predicts that a comparison between units whose variances are either 1.8 or, for instance, 2.8 would show the hypothesized results. The former units (those with smaller receptive fields) would produce a more suitable code for categorical tasks; the latter units (those with larger receptive fields) would produce a more suitable code for coordinate tasks.

The above analysis varied the Gaussian units' receptive field sizes but not the number of units and the units' positions. Consequently, we are unable to ascertain whether the results are due to changes in the size of the units' receptive fields or to changes in the degree of overlap between the units' receptive fields. The work of Hinton (1981) and Ballard (1986) described above strongly suggest that it is the overlap that is most important. If so, then variations in the number of units, the positions of units, or the size of units' receptive fields may all produce similar outcomes. Obviously, it is desirable to understand from a computational viewpoint the consequences of varying all three factors. This article, for methodological reasons, only explores changes in receptive field size and leaves the study of the consequences of varying the other factors as an area for future research.

General Discussion

The computer simulations reported in this article converge in supporting our hypothesis: differences in receptive field sizes can indeed coordinate complementary representations of shape and spatial relations. However, it is worth noting that although the left hemisphere can categorize high spatial frequency gratings faster than the right, and the right can categorize low spatial frequency gratings faster than the left, both hemispheres can *detect* the two sorts of gratings equally well (e.g., see Christman et al., 1991; Kitterle and Selig, 1991). Similarly, the hemispheric differences in classifying local versus global components of hierarchical letters is fragile; it often does not replicate, and a meta-analysis is necessary to detect the effect (see Van Kleeck, 1989). Such results suggest to us that attentional differences may lie at the heart of these hemispheric effects. If so, then our simulations should be interpreted as indicating the virtues of attentional biases in the two hemispheres, not that the hemispheres are hard-wired to encode outputs from neurons with different-sized receptive fields.

In closing, it is worth noting that the experiments presented in this article support two general suppositions about structure-function relationships. First, the nature of a task can influence the structure of a network that learns to perform that task. This was evident in Experiments 2 and 3, where we found that networks' receptive field sizes adapt to the nature of the tasks on which the networks are trained. The second supposition is that the structure of a network can influence the nature of the task that the network learns to perform. This hypothesis finds support in Experiment 1, where we tested networks with different fixed receptive field sizes, and in Experiment 4, where we found that receptive field size is a moderately strong bias in the determination of which expert network learns to perform which task. The fact that receptive field size is not a very powerful bias suggests that in biological systems other factors may also play a role.

Appendix

For completeness, this appendix provides the equations governing learning in the modular architecture. These equations and a fuller discussion of them may be found in Jacobs, Jordan, Nowlan, and Hinton (1991), Jacobs and Jordan (1991), Nowlan and Hinton (1991), and Jordan and Jacobs (1992). The parameters of the expert and gating networks are adjusted simultaneously using the backpropagation algorithm (Rumelhart, Hinton, and Williams, 1986) so as to maximize the objective function

$$\ln L = \ln \sum_{i=1}^{n} \frac{g_i}{\sigma_i}\, e^{-\frac{1}{2\sigma_i^2}||\mathbf{y}^* - \mathbf{y}_i||^2} \tag{5}$$

where $\mathbf{y}^*$ denotes the target vector and σ_i^2 denotes a scaling parameter associated with the i^{th} expert network.

The architecture is perhaps best understood if it is interpreted within a probabilistic framework as an "associative Gaussian mixture model" (see Duda and Hart (1973) and McLachlan and Basford (1988) for a discussion of non-associative Gaussian mixture models).

The training patterns are assumed to be generated by a number of different probabilistic rules. At each time step, a rule is selected with probability g_i and a training pattern is generated by the selected rule. Each rule is of the form $\mathbf{y}^* = f_i(\mathbf{x}) + \epsilon_i$ where $f_i(\mathbf{x})$ is a nonlinear function of the input vector $\mathbf{x}$. If we assume that ϵ_i is a Gaussian random variable with covariance matrix $\sigma^2\mathbf{I}$, then the error vector $\mathbf{y}^* - \mathbf{y}_i$ is also Gaussian and the function in Equation 5 is the log likelihood of generating a particular target vector $\mathbf{y}^*$.

The goal of the architecture is to model the distribution of training patterns. This is achieved by gradient ascent in the log likelihood function. First we consider the partial derivative of the log likelihood with respect to the weighted sum s_i at the i^{th} output unit of the gating network:

$$\frac{\partial \ln L}{\partial s_i} = h_i - g_i \tag{6}$$

where h_i is the a posteriori probability that the i^{th} expert network generates the target vector:

$$h_i = \frac{\dfrac{g_i}{\sigma_i} e^{-\frac{1}{2\sigma_i^2}||\mathbf{y}^* - \mathbf{y}_i||^2}}{\displaystyle\sum_{j=1}^{n} \dfrac{g_j}{\sigma_j} e^{-\frac{1}{2\sigma_j^2}||\mathbf{y}^* - \mathbf{y}_j||^2}}. \tag{7}$$

The weights of the gating network are adjusted so as to minimize the difference between the a posteriori probabilities and the network's outputs, the a priori probabilities g_i.

Consider now the derivatives of the log likelihood with respect to $\mathbf{y}_i$, the output of the i^{th} expert network:

$$\frac{\partial \ln L}{\partial \mathbf{y}_i} = \frac{h_i}{\sigma_i^2}(\mathbf{y}^* - \mathbf{y}_i). \tag{8}$$

These derivatives weight the error term $\mathbf{y}^* - \mathbf{y}_i$ by the a posteriori probability associated with the i^{th} expert network. That is, expert network i's weights are adjusted to reduce the error between its output and the target vector, but only in proportion to its a posteriori probability. Typically only one expert network has a large a posteriori probability for each input vector and, thus, only one expert network learns each training pattern. In general, different expert networks learn different training patterns.

Finally we present the derivative of the log likelihood with respect to the variance σ_i^2 associated with the i^{th} expert network:

$$\frac{\partial \ln L}{\partial \sigma_i^2} = \frac{h_i}{2\sigma_i^4}(\| \mathbf{y}^* - \mathbf{y}_i \|^2 - \sigma_i^2). \tag{9}$$

The variance σ_i^2 is adjusted toward the sample variance $\| \mathbf{y}^* - \mathbf{y}_i \|^2$, but with a step size that is weighted by the a posteriori probability.

References

Ballard, D. H. (1986) Cortical connections and parallel processing: Structure and function. *Behavioral and Brain Sciences*, 9, 67-120.

Ballard, D. H. & Brown, C. M. (1982) *Computer Vision.* Englewood Cliffs, NJ: Prentice-Hall.

Biederman, I. (1987) Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147.

Bridle, J. (1989) Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogelman-Soulie & J. Herault (Eds.), *Neuro-computing: Algorithms, Architectures, and Applications.* New York: Springer-Verlag.

Christman S., Kitterle F., & Hellige J. (1991) Hemispheric asymmetry in the processing of absolute versus relative spatial frequency. *Brain and Cognition*, 16, 62-73.

Dennis, M. & Kohn, B. (1975) Comprehension of syntax in infantile hemiplegics after cerebral hemidecortication: Left-hemisphere superiority. *Brain and Language*, 2, 472-482.

Dennis, M. & Whitaker, H. A. (1976) Language acquisition following hemidecortication: Linguistic superiority of the left over the right hemisphere. *Brain and Language*, 3, 404-433.

Desimone, R. & Ungerleider, L. G. (1989) Neural mechanisms of visual processing in monkeys. In H. Goodglass & A. R. Damasio (Eds.), *Handbook of Neuropsychology.* New York: Elsevier.

Duda, R. O. & Hart, P. E. (1973) *Pattern Classification and Scene Analysis.* New York: John Wiley & Sons.

Hinton, G. E. (1981) Shape representation in parallel systems. In A. Drina (Ed.), *Proceedings of the Seventh International Joint Conference on Artificial Intelligence.*

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986) Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Volume 1: Foundations).* Cambridge, MA: MIT Press.

Hinton, G. E. & Nowlan, S. J. (1990) The bootstrap Widrow-Hoff rule as a cluster-formation algorithm. *Neural Computation*, 2, 355-362.

Jacobs, R. A. & Jordan, M. I. (1991) A competitive modular connectionist architecture. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in Neural Information Processing Systems 3.* San Mateo, CA: Morgan Kaufmann Publishers.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991) Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, 15, 219-250.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991) Adaptive mixtures of local experts. *Neural Computation*, 3, 79-87.

Jordan, M. I. (1986) Serial order: A parallel, distributed processing approach. Technical Report 8604, University of California, San Diego.

Jordan, M. I. & Jacobs, R. A. (1992) Hierarchies of adaptive experts. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in Neural Information Processing Systems 4*. San Mateo, CA: Morgan Kaufmann Publishers.

Kitterle F. & Selig, L (1991) Visual field effects in the discrimination of sine-wave gratings. *Perception & Psychophysics*, 50, 15–18.

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., & Koenig, O. (1992) Categorical versus coordinate spatial representations: Computational analyses and computer simulations. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 562-577.

Kosslyn, S. M. & Koenig, O. (1992). *Wet mind: The new cognitive neuroscience.* New York: The Free Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* New York: W. H. Freeman.

Marsolek, C. J. (1992) Shape representation in the cerebral hemispheres. Unpublished Ph.D. dissertation, Department of Psychology, Harvard University.

Marsolek, C. J., Kosslyn, S. M., & Squire, L. R. (1992) Form-specific visual priming in the right cerebral hemisphere. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 492-508.

McLachlan, G. J. & Basford, K. E. (1988) *Mixture Models: Inference and Applications to Clustering.* New York: Marcel Dekker.

Nowlan, S. J. (1990) Maximum likelihood competitive learning. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2.* San Mateo, CA: Morgan Kaufmann Publishers.

Nowlan, S. J. & Hinton, G. E. (1991) Evaluation of adaptive mixtures of competing experts. In R. P. Lippmann, J. E. Moody, & D. S. Touretzky (Eds.), *Advances in Neural Information Processing Systems 3.* San Mateo, CA: Morgan Kaufmann Publishers.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986) *Numerical Recipes: The Art of Scientific Computing.* New York: Cambridge University Press.

Robertson L. & Lamb M. (1991) Neuropsychological contributions to theories of part/whole organization. *Cognitive Psychology*, 23, 299-330.

Robertson L., Lamb, M., & Knight, R. (1991) Normal global-local analysis in patients with dorsolateral frontal lobe lesions. *Neuropsychologia*, 29, 959-967.

Rumelhart, D. E., Hinton, G., & Williams, R. (1986) Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Volume 1: Foundations).* Cambridge, MA: MIT Press.

Teuber, H. L. (1955) Physiological psychology. *Annual Review of Psychology*, 6, 267-296.

Van Kleeck, M. H. (1989) Hemispheric differences in global versus local processing of hierarchical visual stimuli by normal subjects: New data and a meta-analysis of previous studies. *Neuropsychologia*, 27, 1165-1178.