Encoding Categorical and Coordinate Spatial Relations Without Input-Output Correlations: New Simulation Models

DAVID P. BAKER, CHRISTOPHER F. CHABRIS, AND
STEPHEN M. KOSSLYN

Harvard University

Cook (1995) criticized Kosslyn, Chabris, Marsolek, & Koenig's (1992) network simulation models of spatial relations encoding in part because the absolute position of a stimulus in the input array was correlated with its spatial relation to a landmark; thus, on at least some trials, the networks did not need to compute spatial relations. The network models reported here include larger input arrays, which allow stimuli to appear in a large range of locations with an equal probability of being above or below a "bar," thus eliminating the confound present in earlier models. The results confirm the original hypothesis that as the size of the network's receptive fields increases, performance on a coordinate spatial relations task (which requires computing precise, metric distance) will be relatively better than on a categorical spatial relations task (which requires computing above/below relative to a landmark).

I. INTRODUCTION

What computational principles underlie the spatial processing the brain uses to perform visual cognitive tasks? We and others have begun to answer this fundamental question in the past several years through a convergent series of experiments, with both human participants and computer models. The computer models have recently generated criticism for reasons related to purported conceptual confusions and possible statistical artifacts in their design (Cook, 1995; Cook, Früh, & Landis, 1995). We have replied to these criticisms elsewhere (Kosslyn, Chabris, Marsolek, Jacobs, & Koenig, 1995; Kosslyn, Chabris, & Baker,

1995). This article presents new simulation results in response to perhaps the most funda-
mental issue, whether our previous simulations actually provide evidence that the size of
receptive fields selectively influences one type of spatial relations encoding. This is impor-
tant because the issue centers on whether there actually is more than one way to encode
spatial information, and selective effects in the simulation models are one source of evi-
dence for this distinction.

Specifically, Kosslyn and colleagues (Kosslyn, 1987; Kosslyn, Chabris, Marsolek, &
Koenig, 1992; Kosslyn, Koenig, Barrett, Cave, Tang, & Gabrieli, 1989) postulated the
existence of two distinct subsystems within the visual-spatial processing pathway (the dor-
sal system described by Ungerleider & Mishkin, 1982). One subsystem encodes *categori-
cal* spatial relations, which describe the location, orientation, or other spatial characteristics
of one object relative to another in terms of a broad equivalence class (e.g., A is *above/or
below* B). The other subsystem encodes *coordinate* spatial relations, which specify an
object's spatial attributes with respect to another object in terms of precise metric coordi-
nates (see Kosslyn, 1994). Categorical spatial relations are useful for recognizing and iden-
tifying natural objects that can be seen in many distinct configurations, because categorical
spatial relations tend to remain constant under natural flexion and movement: although an
arm can be in many specific places relative to the trunk of a body, it is always *attached*, and
this information helps one to distinguish human bodies from other objects. As such, cate-
gorical relations tend not to discriminate finely on the basis of metric spatial location but
are instead defined on somewhat broad regions of space; for example, there are many dis-
tinct spatial locations *above* a given object. In contrast, knowing that someone's hand is
attached to her arm is not enough to be able to shake it; you need to know exactly where in
space to direct your hand. Coordinate spatial relations need to refer to specific spatial loca-
tions and distances, and thus do not group different points in space into one equivalence
class.

Kosslyn (1987, 1994) reviewed neuropsychological and behavioral data suggesting that
the left hemisphere is specialized to encode categorical spatial relations more readily than
the right, but that the right hemisphere is specialized to encode coordinate spatial relations
more readily than the left—although each hemisphere is able to process both kinds of
information. He proposed that a small initial imbalance in the processing characteristics of
the two hemispheres' computational systems, perhaps related to innate left-hemisphere
competence for language and right-hemisphere superiority in mapping space, may cause a
"snowball effect" of increasing divergence of neural systems during development, leading
ultimately to the dissociation between these two types of spatial relations encoding (for a
simulation, see Kosslyn, Sokolov, & Chen, 1989). Kosslyn et al. (1992) further refined this
thesis, suggesting that categorical and coordinate spatial relations processing may occur
differentially in the two hemispheres because the two kinds of computations are incompat-
ible: in order to compute the categorical relations among visual objects, the visual system
needs to ignore precise spatial information, whereas exactly that information is required for
coordinate spatial relations computation.

In order to examine the viability of these hypotheses, Kosslyn et al. (1992) conducted a
number of neural-network simulations designed to model human performance on a pair of
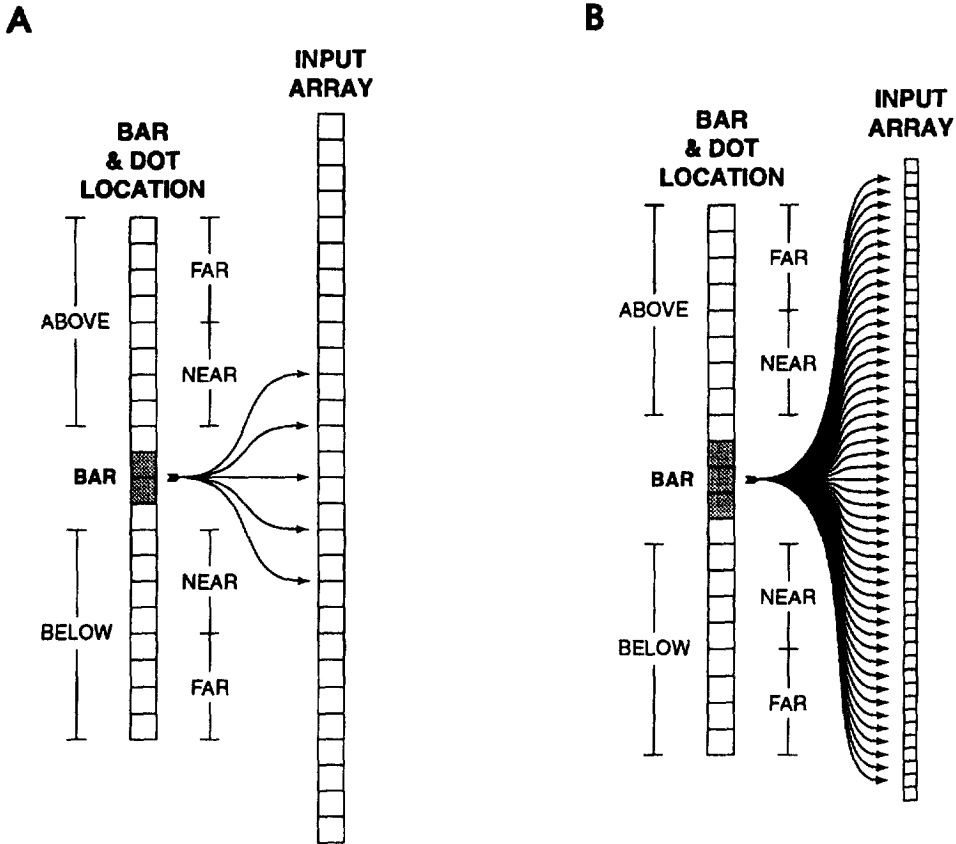
Figure 1A. A schematic diagram of the input patterns for the networks in Kosslyn et al. (1992). The bar consisted of two contiguous activated units, and the dot could be from 2 to 9 units above or below the bar. (Positions fewer than 5 units away from the bar were "near"; others were "far.") The 16 dot-and-bar configurations could be placed in one of five locations in the 28-unit input array, shown by the arrows. (Adapted from Kosslyn et al., 1992.)

Figure 1B. A schematic diagram of the input patterns for the networks used in the current study. The bar consists of three contiguous activated units, and the dot could be from 2 to 9 units above or below the bar. (Positions fewer than 5 units away from the bar were "near"; others were "far.") The 16 dot-and-bar configurations could be placed in any location on the 49-unit input array (so long as all three bar units were within the array). See text for details.

behavioral tasks (Hellige & Michimata, 1989; Kosslyn et al., 1989) that tap into the two kinds of visual-spatial processing. They manipulated several variables in order to test hypotheses about which variables affect the two types of processing differently. More generally, they wanted to use these interactions to confirm the hypothesis that the hemispheric asymmetry is linked to the computational distinction between the two processes.

The neural-network models were trained to perform two tasks, either separately or simultaneously; Figure 1a provides a graphical representation of the tasks. In these models, the stimuli contained a one-unit "dot" that was two to nine units above or below a two-unit "bar," which could occupy one of five positions within the 28-unit, one-dimensional input

space. For the categorical task, the network was trained to respond differentially when the dot was above as opposed to below the bar; for the coordinate task, it was trained to distinguish between stimuli where the dot was close to or far from the bar (within five units or not).

One of the variables examined was receptive field (RF) size. It is known that different arrangements of RFs are appropriate for representing different types of information. In particular, arrays of large, overlapping RFs, although relatively unhelpful individually with regard to spatial location, in aggregate actually provide precise spatial location information. This kind of representation is usually referred to as "coarse coding" (Hinton, McClelland, & Rumelhart, 1986). In contrast, arrays of small, non-overlapping RFs, although they do not encode precise metric information, can be used efficiently to carve the input space into discrete but homogeneous "bins" (Kosslyn et al., 1989). Kosslyn et al. (1992) reasoned that the precise metric information represented by a coarsely coded input would be more useful for computing coordinate spatial relations representations, whereas the discrete delineations produced by units with small RFs would be more useful for computing categorical spatial relations representations. To test that hypothesis, in Part 2 of Study 3 they built networks with large and small hard-wired RFs, which allowed them to examine the interaction between RF size and task type. They found that although the categorical network's performance degraded slightly with the larger RFs, the coordinate network performed much better with them. In addition (in Part 1 of Study 3), networks that were not pre-configured in this way spontaneously developed larger RFs if they were trained to encode coordinate spatial relations than if they were trained to encode categorical spatial relations. These findings suggested that the computations underlying the two tasks are indeed distinct. Moreover, Kosslyn, Anderson, Hillger, and Hamilton (1994) found that the left hemisphere tends to encode information from smaller regions of space than the right, which—in conjunction with the simulation results—suggests that this variable may play a key role in the hemisphere differences observed previously (for a review, see Kosslyn, 1994).

Jacobs and Kosslyn (1994) pursued this line of research further and reported the results of several neural-network simulations examining the interaction of RF size and the categorical/coordinate distinction. When they examined the relative performance of networks on tasks similar to those described above (but with two-dimensional input arrays and more complex stimuli), they too found that the networks improved on the coordinate task as RF size increased, and networks spontaneously developed large RFs if trained on the coordinate task but small RFs if trained on the categorical task.

Crucial predictions derived from the models reviewed above have been confirmed by two subsequent divided-visual-field studies. Cowin and Hellige (1994) asked subjects to make above/below or near/far (greater or less than 3 mm) judgments using the dot-and-bar stimuli developed by Hellige & Michimata (1989), also used by Kosslyn et al. (1989). In one condition the stimuli were presented normally and in another were slightly blurred, effectively removing very high spatial frequencies. Perhaps counterintuitively, blurring impaired categorical but not coordinate judgments. However, if categorical spatial relations are computed using small receptive fields that are sensitive to high frequency infor-

mation, and this information was degraded in the blurred stimuli, then the result is exactly what would be predicted from the Kosslyn et al. (1992) models. Furthermore, if coordinate spatial relations rely on larger receptive fields that are sensitive to low frequency information, and this information was left intact by the blurring, then that manipulation should have no effect on coordinate processing. This is exactly what Cowin and Hellige found.

Roth and Hellige (1997) used the "modified version" spatial judgment tasks of Rybash and Hoyer (1992) and added color manipulations. In their first experiment the stimuli were either green on a subjectively isoluminant red background or were red on green. The former display, with red the dominant color, should selectively impede the magnocellular pathway (e.g., see Breitmeyer & Williams, 1990; Dreher, Fukuda, & Rodieck, 1976), which is associated with low spatial frequencies and hence may rely on outputs from cells with relatively large receptive fields. Roth and Hellige found that the red background slowed coordinate processing, whereas the green background slowed categorical processing. These findings are exactly as predicted by the Kosslyn et al. (1992) models. In Roth and Hellige's second experiment, either the stimuli or the background was held at black while the other varied in color. Now, when red was the only color present, coordinate processing was impaired and categorical processing was not. Taken together, the results of these studies confirm through independent manipulation of spatial frequency and processing pathways the major theoretical claim of Kosslyn et al.—that large receptive fields are more useful than small ones for computing coordinate spatial relations, and vice-versa for categorical spatial relations. If the ability to generate nonobvious (and perhaps counterintuitive) valid predictions of future experimental findings by other laboratories is a criterion for judging scientific theories, then the work of Kosslyn et al. has stood up well so far.

II. INPUT-OUTPUT CORRELATIONS

Cook and his colleagues (Cook, 1995; Cook et al., 1995) have argued that flaws in the design of the networks and the stimuli used in the experiments reported by Kosslyn et al. (1992) and by Jacobs and Kosslyn (1994) may undermine the validity of the conclusions drawn from these simulation results. Many of these criticisms have been rebutted in detail by Kosslyn, Chabris, and Baker (1995) and Kosslyn et al. (1995). Here, we focus on the remaining central criticism, which cannot be rebutted solely by conceptual analysis. It is important to pursue this point because the simulation results not only provide additional evidence for the distinction between the two types of encoding, but also provide testable empirical predictions; if the models themselves are not valid, then they should not be taken seriously enough to be tested. Thus we now present the results of new simulations in an attempt to examine the issue.

Cook et al. (1995) raise the point that because the bar portion of the input patterns used in the neural-network simulations in Kosslyn et al. (1992) could fall in only five locations near the center of the input array, the training set contained "definitive information" that trivialized the computation the network performed. For example, any time one of the upper nine units was active, the dot was above the bar, and therefore the correct response in the categorical task was "above." Cook et al. point out that this kind of disambiguating infor-
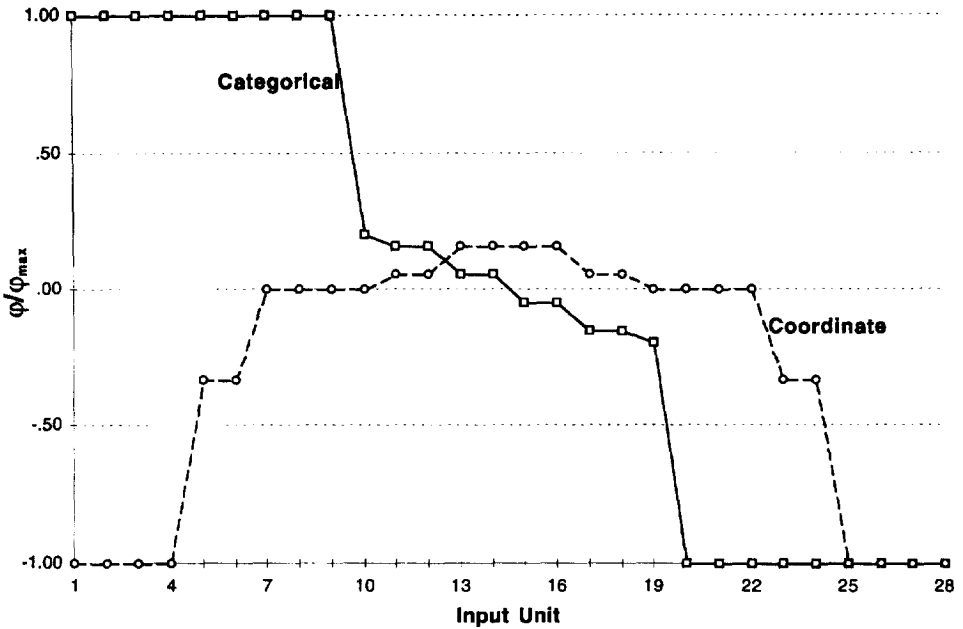
Figure 2. φ/φ_{max} (Cureton, 1959) input-output correlations for each input unit in the Kosslyn et al. (1992) networks, for each task. The magnitude of the average correlation for the categorical task is .69; for the coordinate task it is .36.

mation can be quantified by the φ/φ_{max} correlation coefficient (Cureton, 1959) between the input unit and output unit in question (calculated over all the pattern pairs in the training set), which is exactly 1.0 in such situations. A graph of the φ/φ_{max} input-output correlations for each input unit and the output units that represent each task is shown in Figure 2.

Cook et al. (1995) argue that because some of these correlations are 1.0 and many of the others are high, much of the networks' performance can be attributed to these trivial first-order input-output correlations, and the various results discussed by Kosslyn et al. (1992) cannot with confidence be interpreted as relevant to processes that compute spatial relations representations. In other words, because of the specific training set used, the networks may have learned by "cheating" rather than by developing a computational structure appropriate to the task that humans perform, and so "conclusions of interest to human psychology cannot be drawn" from examining these networks (Cook et al., 1995, p. 420; Cook, 1995, makes a similar point in reference to Jacobs & Kosslyn, 1994).

The results from networks with wired-in RFs are also tainted, Cook et al. point out, because even though filtering the input reduces the correlations below 1.0, it does not make them zero. Cook et al. write, "Because a priori logical analysis is more difficult after receptive fields have been included, it is perhaps best to confirm the absence of input-output imbalance using [φ/φ_{max}] correlation coefficients before implementing receptive fields. Otherwise, effects that are likely due to the manipulation of artifacts may be attributed to changes in receptive field size" (1995, p. 422).

Resolving the Issue

The degree to which a model of a brain system is useful depends upon the degree to which its structure and its input and output representations model those used by the corresponding brain system (for discussions relevant to the present situation, see Kosslyn, Chabris, & Baker, 1995; Kosslyn et al., 1995). In this case, it is arguable (a) whether the high input-output correlations present in the training set account for all of the performance of the system (but if not, how much of each of the various effects examined by Kosslyn et al. was attributable to the "trivial" part of the network's computation, and how much to the nontrivial?), and (b) whether such definitive information is actually unavailable to that part of the brain dedicated to processing this information (Kosslyn et al., 1995, suggested that if the input to the categorical system is a representation that has already been "centered" around a reference point by attentional processes, then the brain may indeed be doing something like the "trivial" computation Cook et al., 1995, identify).

However, before trying to answer such questions, it is important to discover just how important the confounds were in the previous models. If they did not in fact critically affect the networks' performance, then there is no need to speculate further. Thus, in order to remove the confounds in Kosslyn et al.'s (1992) studies, we attempted in Part 1 of this study to replicate the findings of their Study 3, Part 2 using a new training set, a portion of which had zero input-output correlations. We hypothesized that if RF differences, rather than input-output correlations, were truly behind the categorical/coordinate dissociation, then the key effects found by Kosslyn et al. would recur under these more stringent conditions. In Part 2, we examined the generalization ability of the networks, under the assumption that if the network were merely taking advantage of statistical artifacts in the data set, then it would not be able to generalize to instances of the task upon which it had not specifically been trained. The absence of input-output correlations along with demonstrable generalization ability should satisfy even the most ardent critic that conclusions of interest can indeed be drawn from these simulations.

III. SIMULATIONS

PART 1: Removing the Input-Output Correlations

Method

Architecture of the Networks.   The new networks had a feed-forward architecture, with 49 units in the retinal layer, 25 in the RF layer, 15 in the hidden layer, and 2 in the output layer. Each layer of units was fully connected to the next. The connections between the retinal layer and the RF layer were fixed. The weights on the connections feeding into each unit in the RF layer constituted a Gaussian-shaped receptive field:

$$w_{ij} = k(e^{-[(j-j')^2]/\sigma^2})$$

where w_{ij} is the weight from unit j in the retinal layer to unit i in the RF layer, k is a scaling constant, which we set to 5, j' is the index of the input unit at the center of this RF unit's receptive field (the RF units were centered on every other input unit: 1, 3, 5, etc.), and σ is a modifiable term controlling the size of the receptive fields.

All other weight values were set initially to random values uniformly distributed between −0.5 and +0.5 and were modified by the backpropagation learning rule (Rumelhart, Hinton, & Williams, 1986), with the learning rate set to 0.002 and momentum to 0.9.

Training Set.   As illustrated in Figure 1b, each input pattern consisted of a "bar" (three contiguous active units) and a "dot" (one active unit). The bar was allowed to range over the entire input array, and for each position of the bar, the dot ranged from two to nine units above or below the bar. The four positions from 2 to 5 units away from the bar on either side were considered "near"; the four positions greater than 5 units away from the bar were "far."

Those stimuli in which the dot fell off the top or bottom of the input array were discarded, which meant more "far" stimuli were discarded than "near" ones. In order to avoid this discrepancy, for every "far" stimulus that was discarded, a "near" stimulus with the same bar location was discarded as well (specifically, if the dot in the "far" stimulus was n units away from the bar, then we discarded the "near" stimulus whose dot was n–4 units away from the bar). This process produced 632 training patterns (47 bar locations – 16 dot
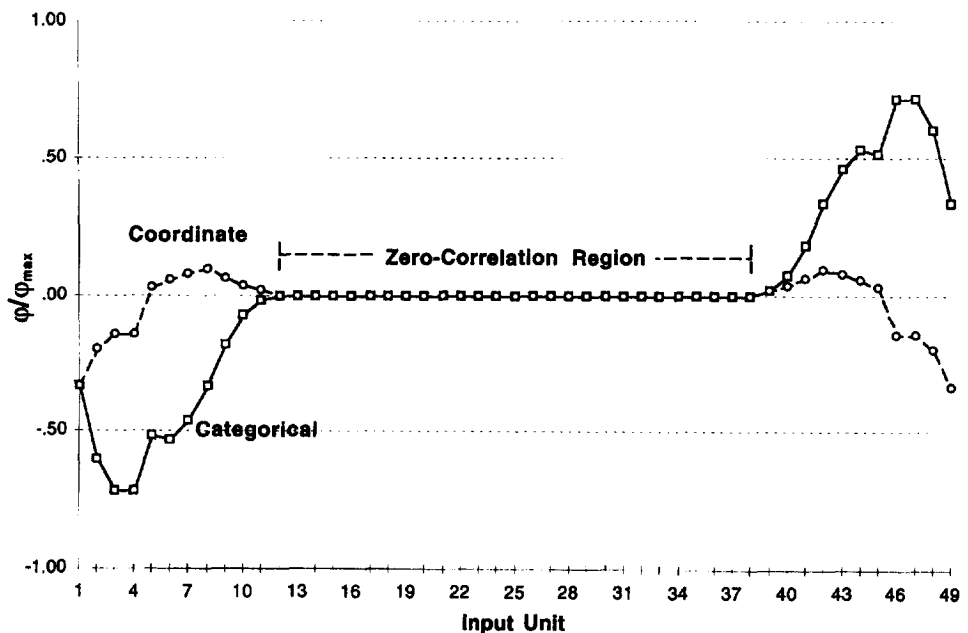

Figure 3. φ/φ_{max} (Cureton, 1959) input-output correlations for each input unit in the networks used in Part 1 of this study, for each task. The magnitude of the average correlation for the categorical task is .18; for the coordinate task it is .02.

locations = 752 initial patterns; 752 − 88 patterns removed after falling off = 664; 664 − 32 more patterns removed to compensate for the near/far discrepancy = 632). The target patterns for the output units were always (0.9 0.1) or (0.1 0.9)—the former if the correct answer was "above" or "near", the latter if the correct answer was "below" or "far."

Figure 3 shows the φ/φ_{max} correlations between each of the 49 input units and the outputs in the two tasks (the nonzero values at the edges of the graph are unavoidable effects of the edges of the input array). These values are much lower than those in Figure 2, which are the analogous data for the training set used in Kosslyn et al. (1992). These low values confirm that this training set does not suffer from the high input-output correlations that the other did. Crucially, note also that there is a central region of 26 units where the input-output correlations are exactly zero for both tasks. This zero-correlation (ZC) region was analyzed separately, as described below.

Procedure. Six levels of RF size (σ) were used: 0.1, 0.5, 1.0, 2.0, 3.0, and 4.0. For each level, the network was trained on each task separately for 2,000 epochs. Two measures of performance were recorded. One was the average squared error (ASE) at the output units (averaged over the two output units and the 632 training patterns), and the other was the fraction of trials for which the network generated an incorrect response (error rate, or ER). A response was considered correct if the output unit that was supposed to be on was greater than the output unit that was supposed to be off. For each of the 12 conditions, five separate simulations were conducted, starting from five different sets of initial weights. The same five sets of weights were used in all 12 conditions so that the data could be treated as a within–initial-weight-set design, analogous to a within-subjects design, for the purposes of statistical analysis.

In addition, we computed separately the ASE and ER for the network's performance on just those 160 stimuli whose four active units were all located within the central ZC region, in order to test whether the network's performance on those portions of its input space that contained no definitive information differed from that on the rest of its input space.

Analyses. We recorded the ASE and ER before learning and after every 10 epochs. As a measure of the network's learning, we used both the average ASE and ER over all 2000 epochs (ASE_a and ER_a), sampled every 10 epochs, as well as the more common measure of ASE and ER at the final epoch (ASE_f and ER_f), after learning had approached asymptote. The average measures, unlike the final measures, have the advantage of taking both the network's ultimate performance and its speed of learning into account, since those networks that reach the same final performance level but at different speeds will have different values of ASE_a and ER_a, as will networks that learn at the same rate but to different extents. We performed a two-factor (task × RF size) ANOVA on each performance value.

To test our hypothesis that the coordinate task would have an increasing advantage over the categorical task as RF size increased, we subtracted the error measures for the coordinate task from those for the categorical task to obtain measures of the advantage of the coordinate over the categorical task for each value of σ, and we then performed a linear contrast on those values.
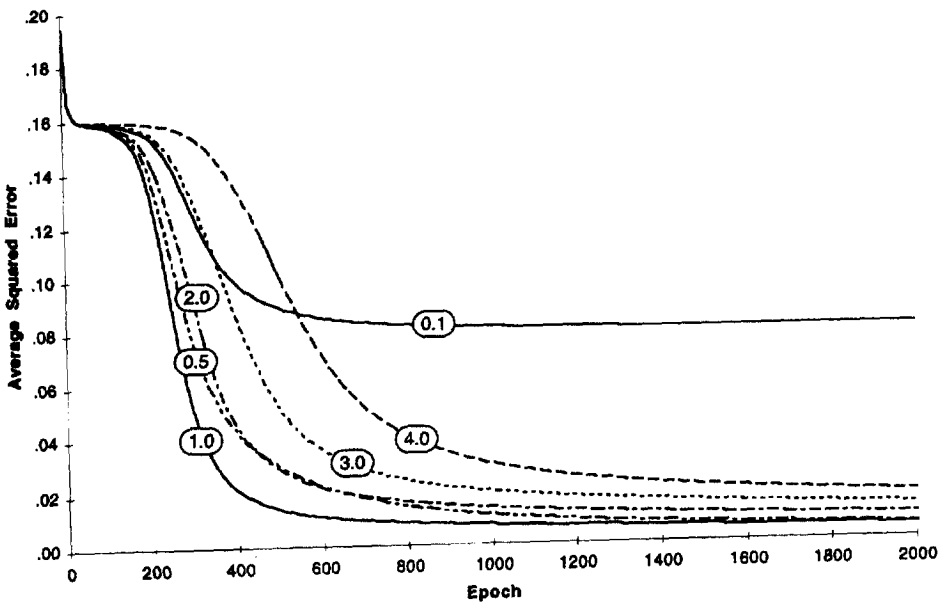
Categorical
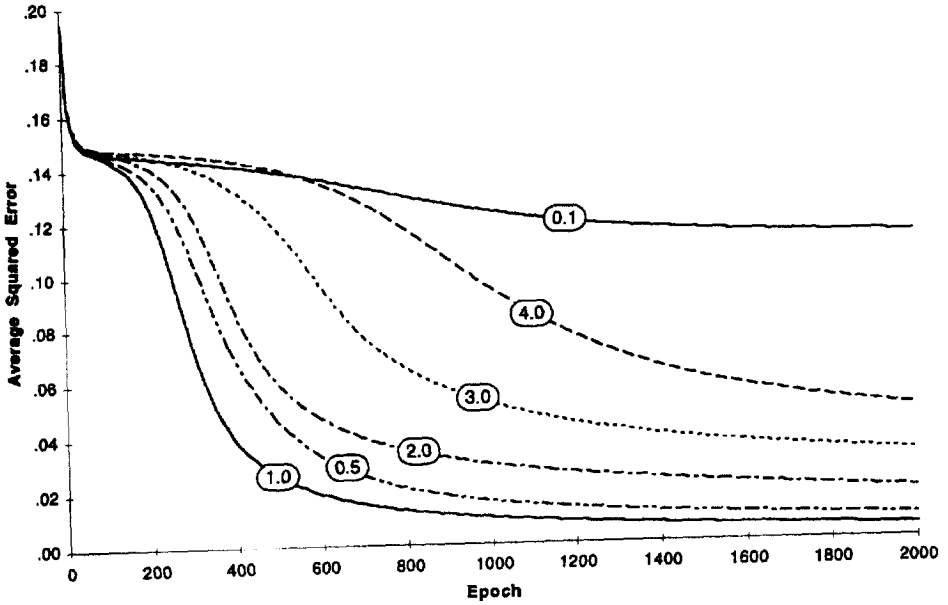
Coordinate

Figure 4. Average squared error on the full training set shown over 2,000 epochs of training in Part 1 of this study. Curves for the categorical task are in the top graph; the bottom graph is for the coordinate task. RF size ranged from $\sigma = 0.1$ to 4.0. See text for details.
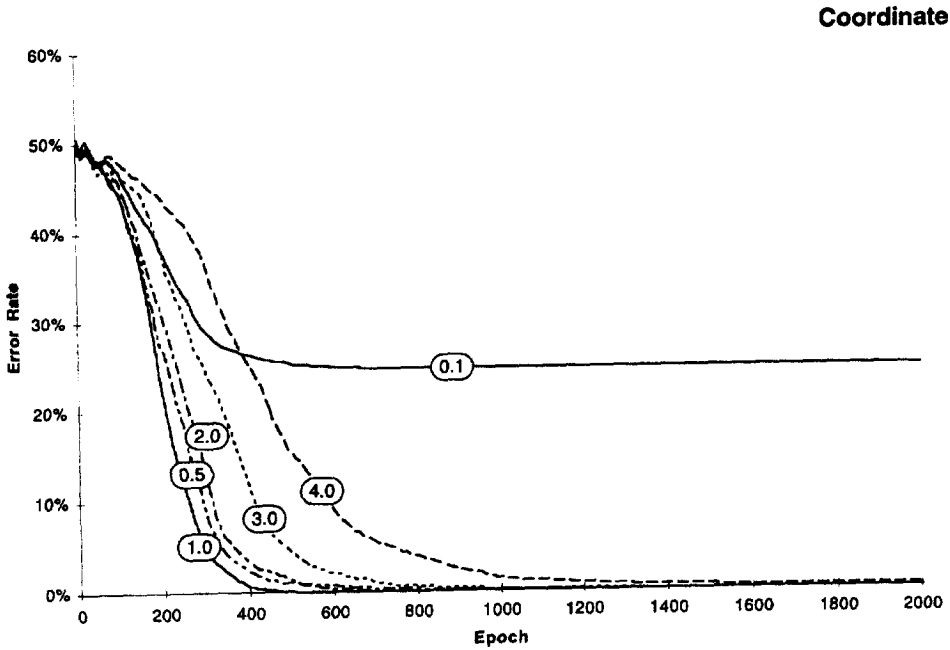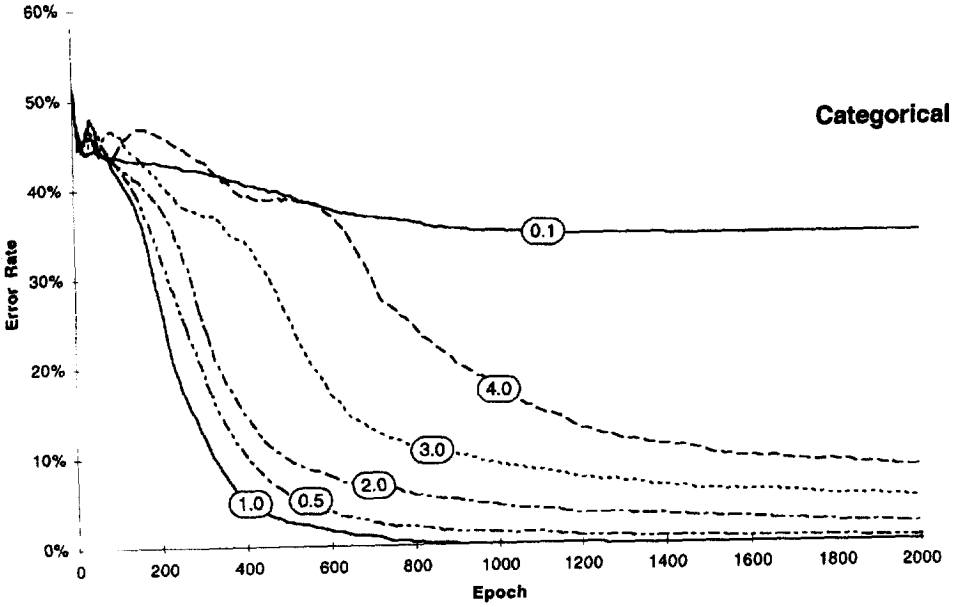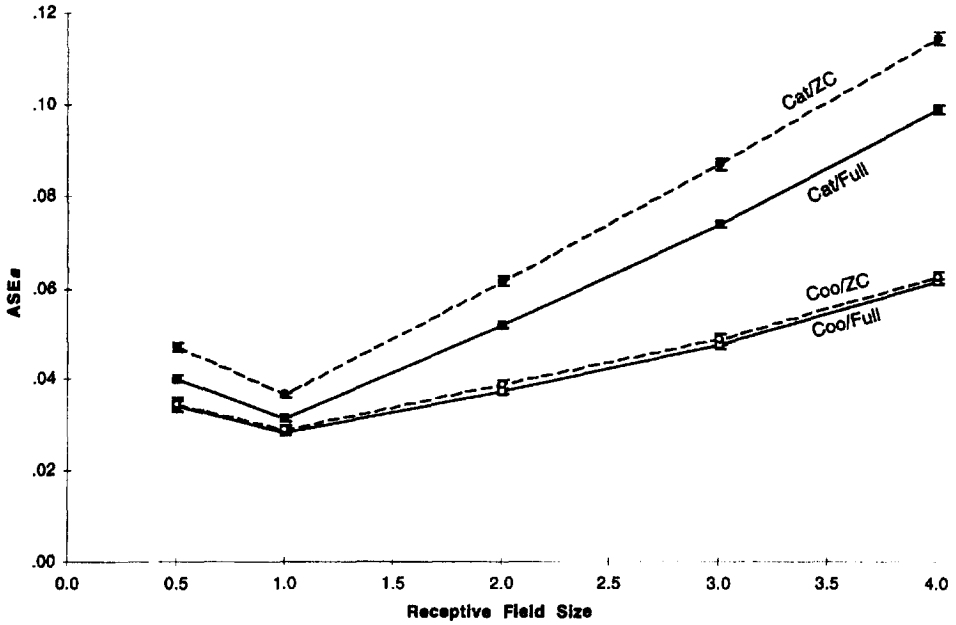
Figure 5. Error rate on the full training set shown over 2,000 epochs of training in Part 1 of this study. Curves for the categorical task are in the top graph; the bottom graph is for the coordinate task. RF size ranged from σ = 0.1 to 4.0. See text for details.

Figure 6. ASE_a on the entire training set ("Full," solid lines) and the zero-correlation subset ("ZC," dashed lines) for the categorical ("Cat," filled dots) and coordinate ("Coo," open dots) tasks at each level of RF size.

Results and Discussion

Figure 4 shows the ASE and Figure 5 shows the ER measures for each task at each RF size. All of the values of σ except 0.1 produced networks that reached an acceptable level of performance. For σ = 0.1, the categorical task's ASE was never lower than 0.11, and its error rate was never lower than 0.34. The coordinate task's ASE never went below 0.08, nor its error rate below 0.24. For all other values of σ, the ASE went below or far below 0.05, and the error rate went below or far below 0.09, which means that the networks were performing, by the end, at over 90% correct. In fact, in many conditions, the networks were at the 100% correct level. The network's poor learning with σ = 0.1 is clearly a result of the Gaussian RFs being too small for the 25 RF units to encode enough information to solve the task. Because of this, the 0.1 level of RF size was not included in any further analyses.

Figure 6 shows the ASE_a for each task for the crucial ZC subset (as well as the entire training set), and Figure 7 shows the difference between the categorical and the coordinate tasks' ASE_a for each value of σ. (Because the patterns for ER_a are so similar, they are not graphed separately here.) The linear contrast on this difference, which is the most direct test of the hypothesis that the coordinate task benefits relatively more from larger RFs than does the categorical task, was highly significant, $F(1,16) = 410$, $p < .0001$ for ASE_a, $F(1,16) = 381$, $p < .0001$ for ER_a. The value of σ affected the ZC subset, $F(4,16) = 1505$, $p < .0001$ for ASE_a, $F(4,16) = 265$, $p < .0001$ for ER_a. In general, the categorical task was harder than the coordinate task, $F(1,16) = 467$, $p < .0001$ for ASE_a, $F(1,16) = 166$,
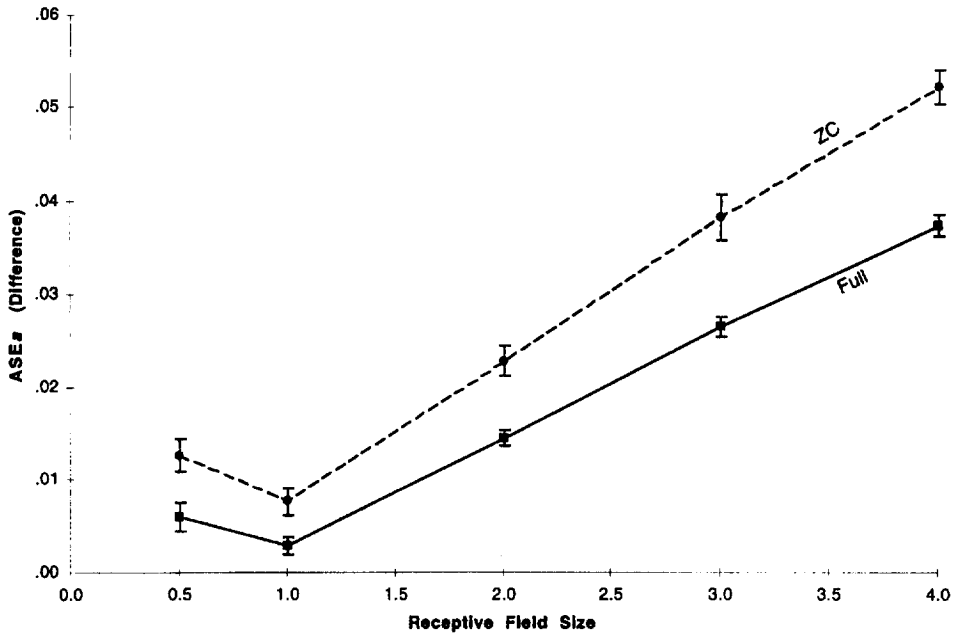
Figure 7. For each level of RF size, the difference between categorical and coordinate (cat–coo) ASE_a (average squared error, averaged over 2000 epochs of training) on the entire training set ("Full," solid line) and the zero-correlation subset ("ZC," dashed line).

$p < .0001$ for ER_a. Finally, there was an interaction between the two, $F(4,16) = 139$, $p < .0001$ for ASE_a, $F(4,16) = 120$, $p < .0001$ for ER_a. The exact same pattern of results was obtained when the entire set of data was examined, $p < .001$ for all comparisons. Moreover, the same pattern of results was found by measuring ASE_f and ER_f in all cases, so we refrain from reporting them separately here.

As hypothesized, the linear contrast in RF size for *categorical–coordinate* performance accounted for significant amounts of variance, which provides strong evidence that increasing RF size is relatively better for processing the coordinate than the categorical task. As Figure 5 shows, although both categorical and coordinate task performance actually decreased over the five values of σ, categorical task performance decreased more. There were also significant differences for both the zero-correlation subset and the entire training set between RF sizes and between tasks, as well as a significant interaction between the two.

The evidence thus argues strongly in favor of our hypothesis. Critically, the fact that the network's performance on the stimuli that fell within the ZC region was the same as on the entire stimulus set (on average, the ASE_a for the ZC region was only 0.0054 higher, and the ER_a was only 1.3% higher, than for the whole stimulus set) indicates that the network did not depend on any "definitive information" of the type that concerned Cook and his colleagues. That the effects, including the crucial linear contrast, for the ZC subset of the training set were just as strong indicates that these more subtle properties of the networks,

and thus of the computation they performed, were also not dependent on any definitive information, since there was none within the ZC region.

PART 2: Generalization Studies

We have assumed that our networks actually computed spatial relations representations. However, in order to make such an assumption with confidence, we must rule out the trivializing possibility that the learning algorithm caused the network to develop the equivalent of a simple table lookup—just "memorizing" each individual stimulus-target pair without performing any true representation-space mapping from input to output. One way to do this is by using a technique in which a randomly selected subset of the stimuli is not used for training the system but is instead reserved for testing the system as training progresses. If the system is learning the underlying computation necessary to transform its input into its output in general, then its performance on "novel" stimuli should also improve as training progresses. However, if performance on novel stimuli starts to worsen even as performance on training stimuli continues to improve, the network has started to sacrifice generalizability for a "rote memorization" procedure, which confers an immediate performance benefit but does not represent the kind of computation we wish to study. Note that rote memorization is a different, more primitive strategy than using sets of small, local RFs to map space into areas or "bins" that correspond to particular spatial relations. The latter strategy would be able to generalize to new trials that fell within the learned areas, whereas rote memorization would not. In this part of the study, we examined the generalization of the networks as they learned the two tasks in order to discover whether the networks are using a trivial strategy to perform their tasks.

Method

Architecture of the Networks.   The architecture of the networks and the tasks were identical to those in Part 1.

Training Set.   The 632 training patterns were split randomly (85% or 536 for training, and 15% or 96 for testing) five separate times, creating five separate generalization trials.

Procedure.   The network models were tested five times on each level of σ (excluding 0.1) and each task from Part 1, varying the random initial weight patterns with the generalization stimulus sets.

Analysis.   ASE and ER values (averaged over the output units and, separately, the 536 training patterns and the 96 testing patterns) were recorded every 10 epochs for 2,000 epochs of training. As a measure of the generalization ability of the network, for each task and level of σ, we found the epoch at which the ASE on the testing set reached a global minimum. We performed the linear *categorical–coordinate* contrast described above using the ASE (and ER) of the training set at the minimum epochs for each task on that run (called minASE and minER). (Again, all analyses were done without the 0.1 level of σ.)
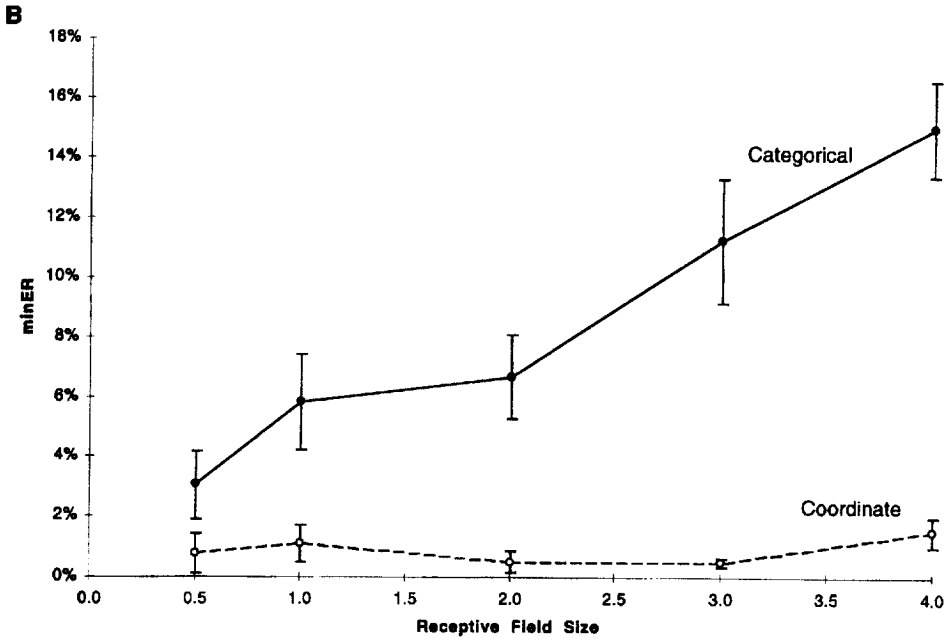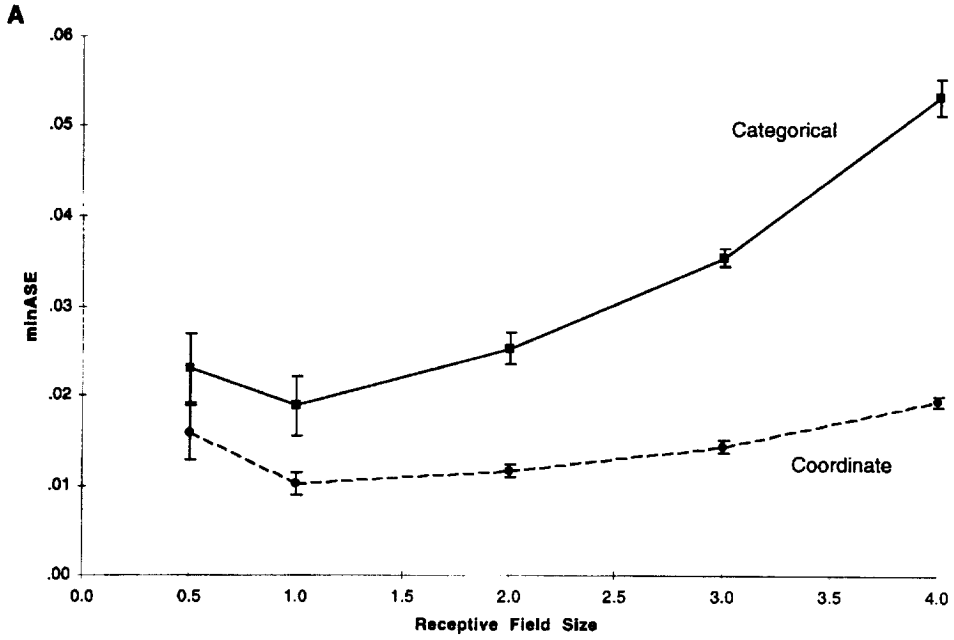
Figure 8. minASE (A) and minER (B) on the testing set in Part 2 of this study.
Curves for the categorical task are solid; dashed curves are for the coordinate task.
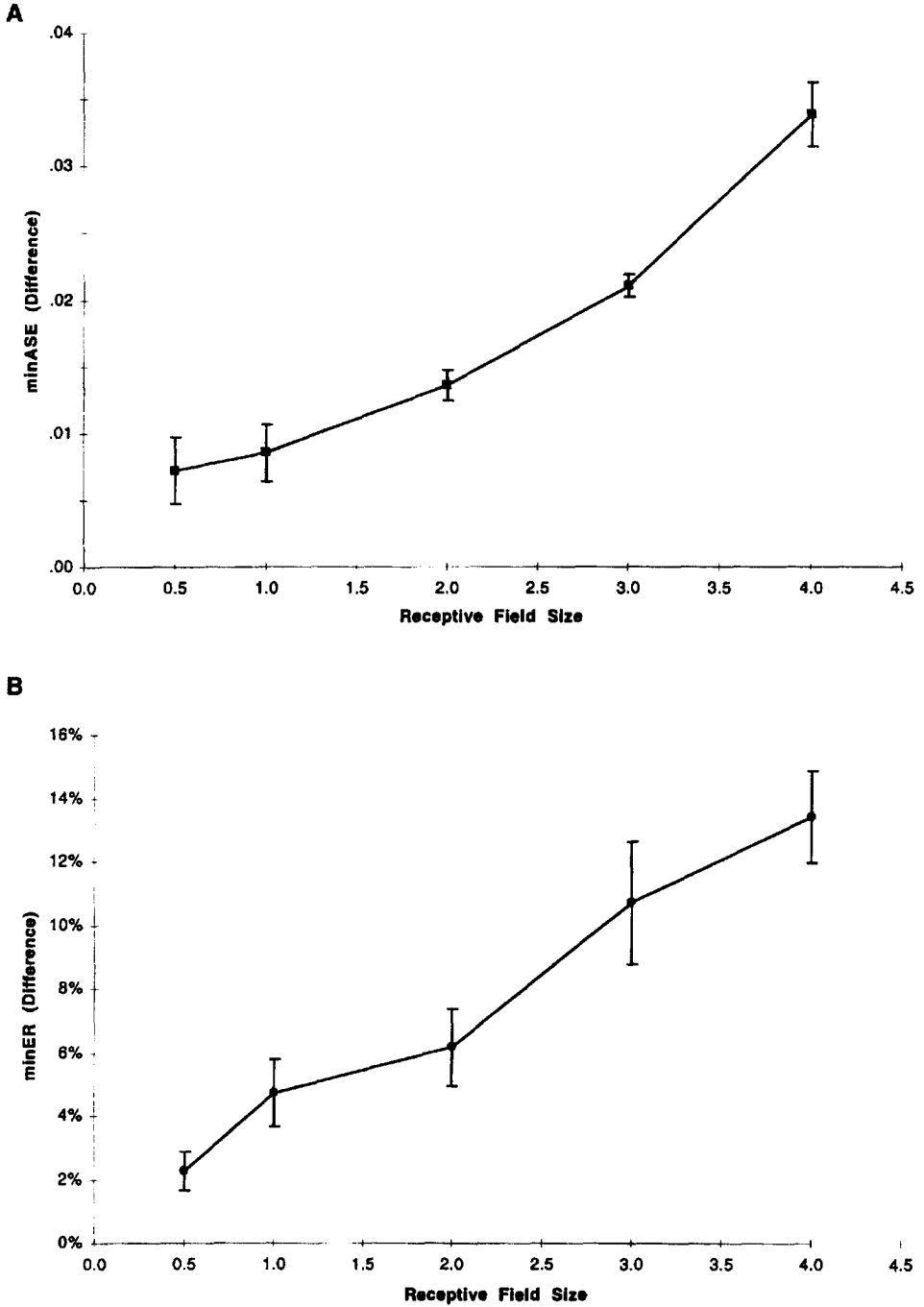RF size ranged from σ = 0.5 to 4.0. See text for details.

Figure 9. For each level of RF size, the difference between categorical and coordinate (*cat–coo*) minASE (A) and minER (B). See text for details.

We used minASE and minER, instead of ASE_a and ER_a, for the generalization trials because ASE_a and ER_a cannot be compared between conditions with different numbers of epochs (the two tasks being compared may reach minima at different times).

Results and Discussion

Figure 8 shows the minASE (8a) and minER (8b), that is, the values after the last of the 2,000 epochs, of the network in the categorical and coordinate tasks, for each level of σ. The network ends up performing at an error rate of just under 9% in the worst condition (categorical at $\sigma = 4.0$), which means that even in the generalization trials, the network is performing at a reasonable level.

Figure 9 shows the average difference (across generalization trials) between the categorical and the coordinate minASE and minER for each RF size. The linear *categorical–coordinate* contrast is highly significant, $F(1,16) = 79.5, p < .0001$ for minASE, $F(1,16) = 38.8, p < .0001$ for minER. As opposed to Part 1, this time the difference is larger for $\sigma = 1.0$ than for $\sigma = 0.5$, supporting our hypothesis, although that difference is not itself significant, $t(4) = 0.34, p > .75$ for minASE, $t(4) = 2.09, p > .10$ for minER.

These results imply that the networks can generalize to stimuli on which they have not been trained, which would not be possible if they were indeed relying on trivial processes instead of a general mechanism that performs the tasks. In addition, the fact that the essential comparison testing our hypothesis, the linear *categorical–coordinate* contrast, remains significant even when the networks' generalization ability is taken into account, provides even more evidence in favor of our hypothesis. Also, the fact that there is an increase in *categorical–coordinate* from $\sigma = 0.5$ to $\sigma = 1.0$ (although not a significant one) when using the minASE measure implies that its reversal in Part 1 may be due to peculiarities of the training set rather than to the true computational properties of the tasks.

IV. GENERAL DISCUSSION

Cook et al. (1995) point out that the presence of high first-order input-output correlations in the stimulus set used by Kosslyn et al. (1992) calls into question the relevance of their models to true categorical and coordinate spatial processing in the brain (and Cook, 1995, does the same for Jacobs & Kosslyn, 1994). They argue that unless this artifact is removed, none of the conclusions Kosslyn et al. and Jacobs and Kosslyn drew are valid. Here, we have presented new simulations that eliminate the input-output correlations in the stimulus set in order to ascertain whether the previous results were misleading. The stimulus set and the network architecture were altered so that the φ/φ_{max} input-output correlations were entirely eliminated in a critical range of the input region. Even so, the results confirmed our hypothesis that the relative advantage for the coordinate task over the categorical task would increase with increasing receptive field size. Finally, when we examined the generalization ability of the networks directly, using a cross-validation procedure, we found that the networks were able to generalize (except when the receptive field size was too small), and that our prediction held even when we considered only the networks' generalizable performance.

We note, however, that the results of these simulations differed from the results of Kosslyn et al. (1992, Study 3, Part 2) in two ways. First, performance on both tasks worsened as receptive field size increased from $\sigma = 1.0$ to 4.0 (see Figure 5). This may be explained by the interaction between RF size and the numbers of units in the input and RF layers, both of which were fixed in this study. Clearly as RF size grows it should eventually reach a point at which neither task can be performed. It is possible that for a given input space size and task, there is an optimal overlap for the representation created by the RFs, and that our range of RF sizes was not able to capture it. Since the corresponding parameters for the relevant brain areas are not currently known, complete resolution of this issue awaits further research.

Second, in the current study network performance on the categorical task was always worse than on the coordinate task (see Figure 5). Kosslyn et al. (1995) argued that "definitive information" was actually available to subjects in the human experiments being modeled by the simulations of Kosslyn et al. (1992), and that excluding it from the networks might actually reduce their value in explaining the phenomena originally under study. Our results support this contention: Comparing our simulations to those of Kosslyn et al. (1992), definitive information was reduced for the categorical task more than for the coordinate task (as measured by φ/φ_{max}: .69 to .18 for categorical, .36 to .02 for coordinate), and the categorical task suffered a relative performance penalty accordingly. And incidentally, the fact that in our tasks the categorical mapping contains higher input-output correlations but is nevertheless more difficult shows that input-output correlations alone cannot explain network performance.

Cook et al. (1995) have reminded us that it is always important, especially when using neural-network simulations, to ensure that one's experimental design can rule out alternative hypotheses. Now that we have ruled out a trivial explanation for the effects Kosslyn et al. (1992) found, we have produced more substantial evidence supporting the computational distinction between processes that encode categorical and coordinate spatial relations representations.

REFERENCES

Breitmeyer, B. B., & Williams, M. C. (1990). Effects of isoluminant-background color on metacontrast and stroboscopic motion: Interactions between sustained (P) and transient (M) channels. *Vision Research, 30,* 1069–1075.

Cook, N. D. (1995). Correlations between input and output units in neural networks. *Cognitive Science, 19,* 563–574.

Cook, N. D., Früh, H., & Landis, T. (1995). The cerebral hemispheres and neural network simulations: Design considerations. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 410–422.

Cowin, E. L., & Hellige, J. B. (1994). Categorical versus coordinate spatial processing: Effects of blurring and hemispheric asymmetry. *Journal of Cognitive Neuroscience, 6,* 156–164.

Cureton, E. E. (1959). Note on Phi/PhiMax. *Psychometrika, 24,* 89–91.

Dreher, B., Fukuda, Y., & Rodieck, R. W. (1976). Identification, classification, and anatomical segregation of cells with X-like and Y-like properties in the lateral geniculate nucleus of old-world primates. *Journal of Physiology (London), 258,* 433–452.

Hellige, J. B., & Michimata, C. (1989). Categorization versus distance: Hemispheric differences for processing spatial information. *Memory & Cognition, 17,* 770–776.

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,* Vol.1: *Foundations.* Cambridge, MA: MIT Press.

Jacobs, R. A., & Kosslyn, S. M. (1994). Encoding shape and spatial relations: The role of receptive field size in coordinating complementary representations. *Cognitive Science, 18,* 361–386.

Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review, 94,* 148–175.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Kosslyn, S. M., Chabris, C. F., & Baker, D. P. (1995). Neural network models as evidence for different types of visual representations. *Cognitive Science, 19,* 575–579.

Kosslyn, S. M., Chabris, C. F., Marsolek, C. J., Jacobs, R. A., & Koenig, O. (1995). On computational evidence for different types of spatial relations encoding: Reply to Cook et al. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 423–431.

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

Kosslyn, S. M., Anderson, A. K., Hillger, L. A., & Hamilton, S. E. (1994). Hemispheric differences in sizes of receptive fields or attentional biases? *Neuropsychology, 8,* 139–147.

Kosslyn, S. M., Koenig, O., Barrett, A., Cave, C. B., Tang, J., & Gabrieli, J. D. E. (1989). Evidence for two types of spatial representations: Hemispheric specialization for categorical and coordinate relations. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 723–735.

Kosslyn, S.M., Sokolov, M.A., & Chen, J.C. (1989). The lateralization of BRIAN: A computational theory and model of visual himispheric specialization. In D. Klahr & K. Kotovsky (Eds.), Complex information processing: The impact of Herbert Simon. Hillsdale, NJ: Erlbaum.

Roth, E. C., & Hellige, J. B. (1998). Spatial processing and hemispheric asymmetry: Contributions of the transient/magnocellular visual system. *Journal of Cognitive Neuroscience, 10,* 472–484.

Rumelhart, D. E., Hinton, G. E., & 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,* Vol.1: *Foundations.* Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition,* Vol. 1: *Foundations.* Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition,* Vol.1: *Foundations.* Cambridge, MA: MIT Press.

Rybash, J.M., & Hoyer, W.J. (1992). Hemispheric specialization for categorical and coordinate spatial representations: A reappraisal. *Memory & Cognition, 20,* 271–276.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *The analysis of visual behavior.* Cambridge, MA: MIT Press.