The neural development and organization of letter recognition: Evidence from functional neuroimaging, computational modeling, and behavioral studies

THAD A. POLK*[†] AND MARTHA J. FARAH[‡]

*Department of Psychology, University of Michigan, 525 E. University, Ann Arbor, MI 48109-1109; and [‡]Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104-6196

ABSTRACT   Although much of the brain's functional organization is genetically predetermined, it appears that some noninnate functions can come to depend on dedicated and segregated neural tissue. In this paper, we describe a series of experiments that have investigated the neural development and organization of one such noninnate function: letter recognition. Functional neuroimaging demonstrates that letter and digit recognition depend on different neural substrates in some literate adults. How could the processing of two stimulus categories that are distinguished solely by cultural conventions become segregated in the brain? One possibility is that correlation-based learning in the brain leads to a spatial organization in cortex that reflects the temporal and spatial clustering of letters with letters in the environment. Simulations confirm that environmental co-occurrence does indeed lead to spatial localization in a neural network that uses correlation-based learning. Furthermore, behavioral studies confirm one critical prediction of this co-occurrence hypothesis, namely, that subjects exposed to a visual environment in which letters and digits occur together rather than separately (postal workers who process letters and digits together in Canadian postal codes) do indeed show less behavioral evidence for segregated letter and digit processing.

Localization of function is a basic feature of brain organization, revealed by the selectivity of impairments after brain damage and by techniques for recording regional brain activity. A wide variety of behavioral functions are now thought to be localized in the brain, ranging from sensorimotor functions such as motor control and sensation in the various modalities to high level cognitive functions such as explicit learning.

What leads such functions to become localized in the brain? For the previous cases, the answer is presumably genetics. Sensorimotor functions, explicit learning, and many other examples of localized functions are old on an evolutionary scale, are shared with other species, provide a clear adaptive advantage, and develop automatically with no systematic training. It is thus possible that the brain has evolved to dedicate tissue to these functions. Consistent with this hypothesis, there is no evidence for the localization of functions like ballet dancing or chess playing: These functions are not old on an evolutionary scale, are not shared with other species, do not provide a clear adaptive advantage, and do not develop automatically without systematic training.

There are, however, a few cognitive functions that may violate this simple generalization. For example, handwriting can be impaired selectively by brain damage while other manual motor control tasks are relatively well preserved (1). Indeed, recent evidence suggests that even the writing of cursive vs. print and uppercase vs. lowercase (2) can dissociate after brain damage. The fact that these functions can be impaired selectively suggests that their neural substrates are localized and spatially segregated from each other. But given that writing is such a recent human accomplishment, that it is not shared by all humans, much less by other species, and that it requires systematic training to develop, a genetic explanation is very unappealing.

Similarly, brain damage can impair selectively the ability to recognize musical melodies without affecting the recognition of other sounds (3). Again, this finding suggests that aspects of musical processing are localized in the brain. But like writing, music is a relatively recent development compared with sensation and motor control, and furthermore it does not provide any obvious adaptive advantage. So once again, a genetic account seems unnatural.

Perhaps most striking of all is evidence for a dissociation between letter and number recognition. In rare cases, patients who have a profound deficit in recognizing letters have significantly less difficulty in recognizing digits and numbers (4). This finding is somewhat ambiguous because of the different sizes of the letter and digit sets. If forced to guess on the basis of partial or uncertain information, the smaller number of possibilities among digits (10 as opposed to 26) will result in better performance. A recent study using electrodes chronically implanted on the surface of striate and extrastriate cortex found sets of neurons that respond to letters but not digits (5). However, the extremely local and widely spaced nature of this recording method makes it difficult to determine whether the letter-sensitive cells are segregated into a functionally defined area. In any case, if letter recognition is indeed localized in the brain, then a genetic account is again problematic: Reading is recent, is not shared with other species, and requires systematic training to develop.

These examples raise a very important question: How could noninnate functions such as these come to depend on dedicated and segregated brain tissue? After all, these functions are in many ways more like chess and ballet dancing than they are like motor control or explicit learning, so their neural substrates presumably could not be hard-wired into our genes. There instead must be certain environmental factors that somehow lead to a change in brain organization such that these functions become localized. What those factors are and how they produce such a change in brain organization are unknown.

In this paper, we will focus on letter and digit recognition and will address these issues by attempting to answer two

[†]To whom reprint requests should be addressed. e-mail: tpolk@ umich.edu.

questions: (*i*) Does the neural architecture of vision include a brain area for letter recognition (as opposed to digit recognition) and (*ii*) if so, how might such an area arise? We will first describe a functional neuroimaging experiment that did indeed find evidence that letter and digit recognition depend on different neural substrates. We will then describe one potential hypothesis that could account for the localization of noninnate functions. Finally, we will present a behavioral study that confirms one critical prediction of that hypothesis with respect to letter and digit recognition.

FUNCTIONAL NEUROIMAGING

In a previous experiment (T.A.P., M. Stallcup, G. K. Aguirre, D. Alsop, M. D'Esposito, J. Detre & M.A.F., unpublished work) we used functional MRI to test whether letter recognition is localized in the brain and, in particular, whether its neural substrate is segregated from that of digit recognition.

Subjects passively viewed blocks of letter strings, blocks of digit strings, blocks of geometric shape strings, and blocks of fixation points (baseline). One subject participated twice (6 weeks apart) for a total of six sessions. The strings of letters, digits, and shapes were matched in length and size, and the letters and digits were presented in the same font. A surface coil was placed over the left occipitotemporal cortex.

Statistically significant segregation was observed in individual subjects. In four of the six sessions, an area in the left inferior occipitotemporal cortex responded significantly more to letters than digits (Fig. 1). Two sessions were run on the same subject (H.B.), 6 weeks apart, and showed activation in the same area both times. The two subjects who did not show significant activation in the letter vs. digit comparison both showed subthreshold activation in the same left inferior occipitotemporal area [K.H. had 17 contiguous voxels above $Z = 2.5$ around the Talairach coordinate $(-33, -34, -4)$; M.S. had 19 contiguous voxels above $Z = 2.5$ around the Talairach coordinate $(-35, -38, -4)$]. The digit vs. letter comparison did not show any significant activations at the $P < 0.016$ level (0.05 after correcting for three planned comparisons) in any subject although this comparison did show one activation at the $P < 0.03$ level (0.1 corrected) in one subject (J.N.). The shape vs. letter/digit comparison showed significant areas of activation in four of the six sessions. Some posterior areas were activated significantly by all three stimulus types relative to fixation. It is also important to point out that using a surface coil over the left hemisphere disrupts the signal in the right hemisphere. One should therefore not conclude that there were no right hemisphere activations or that there were no differences in the comparisons that failed to show significant activations.
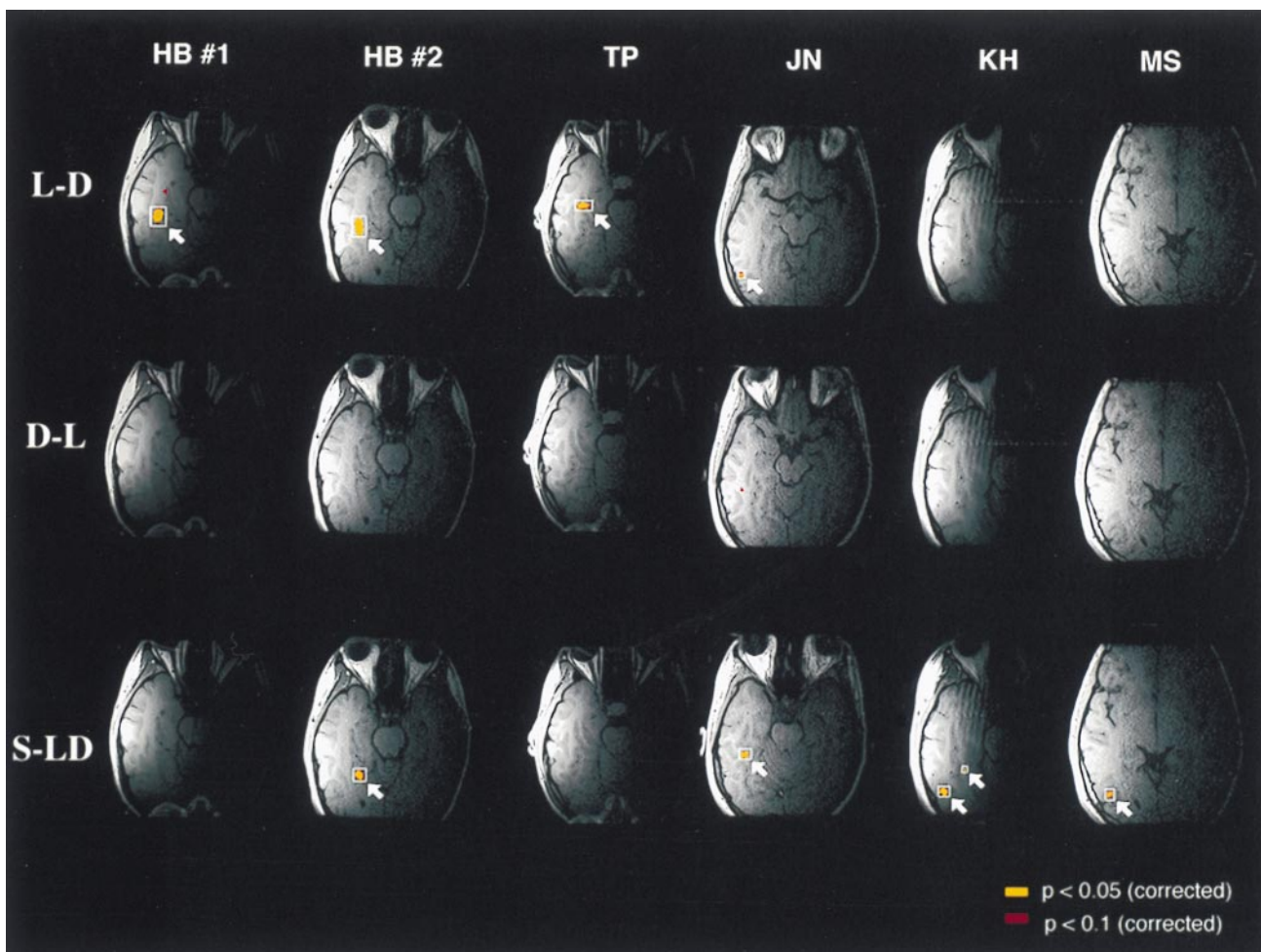

Fig. 1. Significant differences in the blood-oxygenation level-dependent (BOLD) MRI signal during passive viewing of letters vs. digits (L-D), digits vs. letters (D-L), and shapes vs. letters and digits (S-LD) (T.A.P., M. Stallcup, G. K. Aguirre, D. Alsop, M. D'Esposito, J. Detre & M.A.F., unpublished work). For each of the three comparisons in each of the six sessions, the single horizontal brain slice that showed the most significantly activated voxels for that comparison is shown. Voxels in yellow were significant at the *P* < 0.05 level after correcting for all the voxels as well as for the three planned comparisons. Voxels in red were significant at the *P* < 0.1 level, corrected. The left hemisphere appears on the left and the right hemispheres on the right.

These results demonstrate that, at least in some literate subjects, certain extrastriate visual areas respond significantly more to letters than digits and that other areas respond significantly more to shapes than to letters and digits. The fact that certain visual areas responded more to shapes than letters and digits could simply reflect the fact that the shapes looked quite different—they were filled-in polygons rather than sets of lines. Consequently, the shape vs. letter/digit activations may have nothing to do with the stimulus category (shape vs. orthography) but may instead reflect processing of low level visual information (e.g., spatial frequency).

In contrast, there are no obvious physical features that distinguish letters and digits. Indeed, many letter–digit pairs (e.g., O/0, I/1) are much more physically similar than any letter–letter or digit–digit pair. Therefore, the fact that certain brain areas respond significantly more to letters than digits indicates that the visual system processes letters as a category differently than digits and numbers. That is, letter recognition is at least partly segregated from digit recognition.

A natural question is whether this putative "letter area" is actually being activated by letters as opposed to being deactivated by digits. One way to address this question is to compare the processing of letters and digits with the baseline condition (fixation points) in this letter area. In five of the six sessions (the only exception being that of J.N.), we found that this area was much more activated by letters than by fixation and was slightly more activated by digits than fixation. In all five of these sessions, the letter vs. fixation comparison was significant in this area even before restricting attention to the region of interest (that is, even when correcting for all voxels). Although not significant in any session when correcting for all voxels, the digit vs. fixation comparison was significant in three of the six sessions (H.B.1, H.B.2, K.H.) when correcting only for the voxels that were significantly active in the letter vs. digit comparison (for K.H. and M.S. there were no such voxels, so we used voxels for which the Z score of the letter vs. digit comparison was >2.5). In all but one subject, this letter area appears to be activated by letters rather than deactivated by digits.

J.N. showed the opposite pattern. In this subject, the letter vs. fixation comparison did not even approach significance (in either direction) within the area that showed significant letter vs. digit activity, but digits activated this area significantly less than baseline (even when correcting for all voxels). Because these results were so different from the rest of the data (e.g., the letter area is much more lateral, the letter vs. digit activity is due to deactivation by digits, the digit vs. letter comparison approached significance) and because the letter vs. digit activity was near the edge of the brain where motion can easily introduce noise, this subject's results should be interpreted with caution.

The finding of a letter area that is significantly more activated by letters than digits and slightly more activated by digits than fixation can be interpreted in at least two ways. One possibility is that this brain area deals with both letters and digits but is especially active in letter processing. According to this view, this area is specialized for letter processing but is involved in processing digits as well although to a lesser degree. An alternative interpretation is that this brain area is dedicated exclusively to letter processing and only responds to nonletter stimuli to the extent that they are similar to letters; digit strings activate this area more than fixation simply because they are more letter-like. According to this view, this brain area plays no functional role in digit recognition but is nevertheless partially activated by digits.

Under either interpretation, the fact that a brain area responds significantly more to letters than digits has important implications given that letter recognition is not innate: Somehow the environment is producing a significant, qualitative change in brain organization. This is not a case in which experience is modulating the size of a brain area. Rather, experience is leading to the development of a new functional brain area that is especially involved in processing letters. The environment is somehow causing the brain to reallocate a localized part of the visual system to deal with letter recognition—a noninnate function—in a special way.

THE CO-OCCURRENCE HYPOTHESIS

Having established the existence of a letter area, the obvious question is how did it arise? One possible hypothesis is motivated by the observation that neural learning is fundamentally correlation-based. A fundamental insight underlying Hebb's pioneering work on cell assemblies (6) was that neural learning is based on strengthening the connections between simultaneously firing (i.e., correlated) neurons. Put simply, neurons that fire together, wire together.

Given that neural learning is correlation-based, what environmental factors would be most likely to lead to changes in brain organization, such as the emergence of a letter area? One possibility is correlations in the environment. Furthermore, the nature of text and reading imposes some very strong spatial and temporal correlations on letters and words. An obvious characteristic of text is that letters appear together in space. Words are comprised of spatial clusters of letters, sentences are made up of spatial clusters of words, and text is made up of spatial clusters of sentences. These spatial correlations also lead to temporal correlations during reading: Groups of letters are processed together followed immediately by the processing of other groups of letters. Digits, shapes, and other stimuli occur occasionally, but they are the exception; the rule is letters upon letters upon letters.

We hypothesized that this statistical organization of the environment could lead to changes in the spatial organization of the neural architecture underlying visual word recognition, specifically that spatial and temporal clustering of letters could interact with the brain's correlation-based learning mechanisms (Hebbian learning) to lead to the segregation of letter recognition. The feasibility and explicitness of this co-occurrence hypothesis was confirmed by means of a neural network model, and one of its predictions was borne out in a behavioral study, both of which we now describe.

NEURAL NETWORK MODEL

In a previous paper (7), we described a simple 2-layer neural network that uses a Hebbian learning rule to modify the weights of the connections between the input and output layers. Specifically, if two units are both firing (correlated), then their connection is strengthened; if only one unit of a pair is firing (anticorrelated), then their connection is weakened. The input layer represents the visual forms of input characters (letters and digits) by using a localist representation (each unit represents a different visual form). We also constructed a network that used a distributed input representation, but because it produced the same results and because the localist representation makes the operation of the model more transparent, we will focus on the localist network for the rest of this discussion.

Initially, the network's output layer does not represent anything (because the connections from the input layer are initially random), but with training it should self-organize to represent letters and digits in segregated areas. Neighboring units in the output layer were connected via excitatory connections, and units further away were connected via inhibitory connections, in keeping with previous models of cortical self-organization (8–9). (Other architectures would also be consistent with our explanation, e.g., normalization of output activations as opposed to long range inhibitory connections. What is critical is that the architecture provide a cooperative
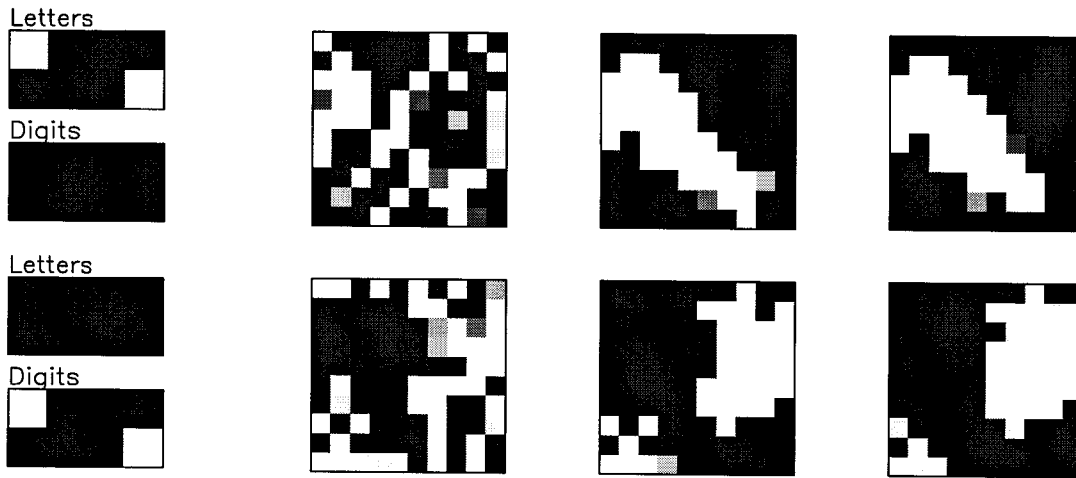
FIG. 2.   The state of the network after 1, 3, and 5 processing cycles when a pair of letters is presented (top row) and when a pair of digits is presented (bottom row) (7). The input layer is on the left, and state of the output layer after 1, 3, and 5 processing cycles is on the right. [Reproduced with permission from ref. 7 (Copyright 1995, *Proceedings of the National Academy of Sciences*)].

mechanism to produce clusters of activity and a competitive mechanism to inhibit multiple clusters. For a review of a variety of such models, see ref. 10. For related models, see refs. 8 and 9 and 11–17.)

Fig. 2 shows the network's behavior when multiple letters are presented simultaneously (top row) as well as when multiple digits are presented simultaneously. The pattern of connectivity in the output layer (neighbors excite, others inhibit) leads to a cluster of activity around the most active output units (18). If two stimuli initially activate widely
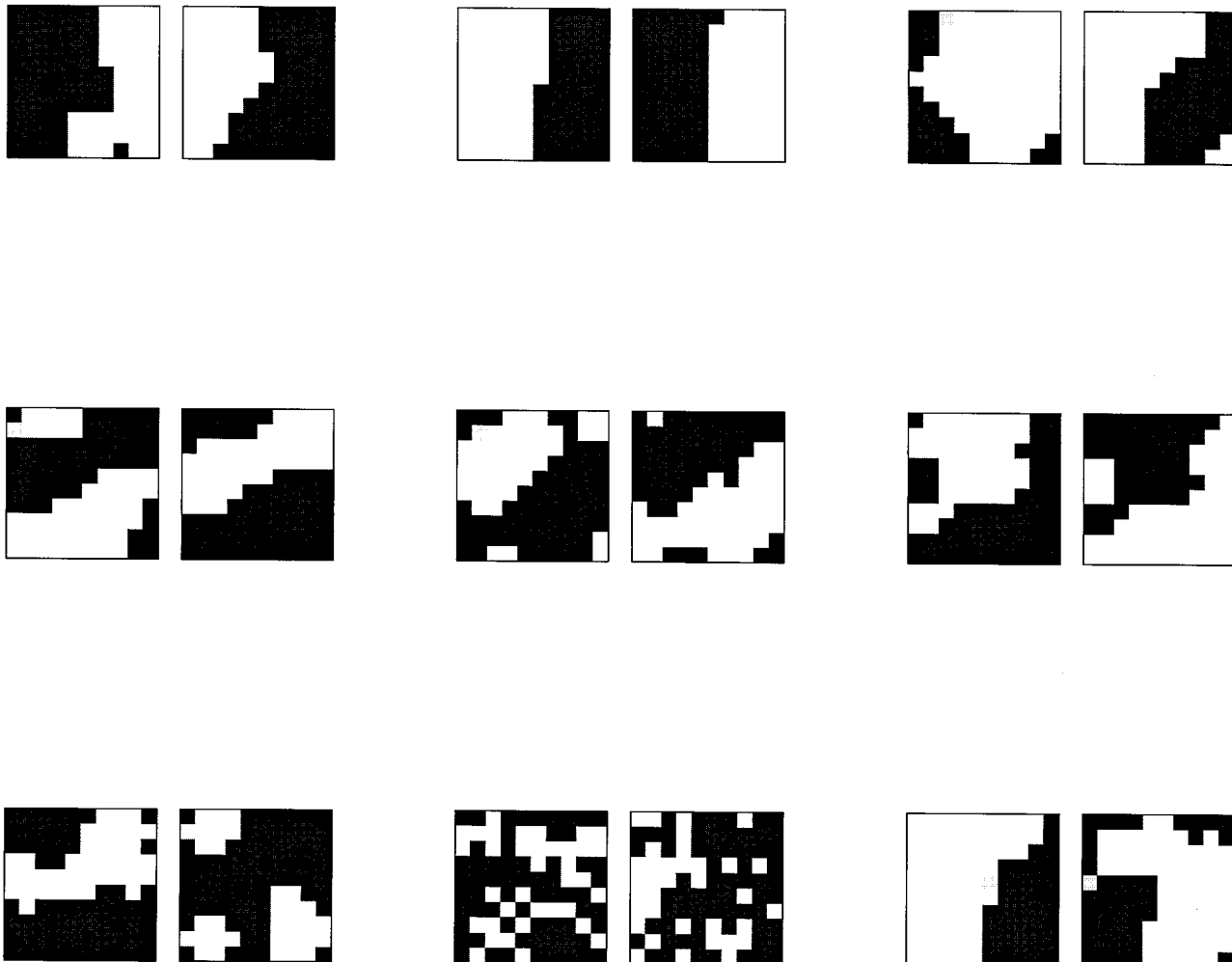

FIG. 3.   The state of the neural network after five epochs of training under varying initial conditions (7). The left of each pair shows the activation when all eight letters are present, and the right shows the activation when all eight digits are present. [Reproduced with permission from ref. 7 (Copyright 1995, *Proceedings of the National Academy of Sciences*)].

separated clusters but then appear together, the initial clusters will compete with each other (via the inhibitory connections) in representing the pair. One cluster will eventually win out, and Hebbian learning will strengthen the connections from both inputs to the victorious cluster. The result is that co-occurring stimuli will be biased toward exciting nearby units, even if they initially excited quite different sets of units. In other words, spatially localized areas will develop for stimuli that tend to co-occur (as we assume letters do). Stimuli that occur in rapid succession also could become associated assuming some residual activation from the first stimulus (19).

The same argument implies that stimuli that do not co-occur will be biased away from exciting nearby units. When one stimulus is present, the connections from any other (inactive) stimulus to the currently active output units will be weakened, and this will bias the inactive stimulus away from exciting those units. Consequently, stimuli from different categories (e.g., letters vs. digits) will tend to be represented by spatially segregated sets of units (because we assume that they co-occur much less frequently). And even within a category, as long as particular stimuli do not always co-occur, they will have distinct representations within their cortical areas.

We trained this network by using clusters of letters and clusters of digits but not with clusters that contained both (to simulate the co-occurrence statistics that were hypothesized to be critical). Fig. 3 shows the results with different initial conditions. Distinct, spatially localized letter and digit areas developed in almost every case. The parameters affected the size, coherence, and degree of overlap of clusters but did not change the qualitative pattern of results. Also note that the resulting letter and digit areas did not always form in the same locations. In some simulations, the letter area arose on the left of the output layer with the digit area on the right. In others, these locations were reversed, or horizontal or diagonal patterns arose.

Letters occur far more frequently than digits, and the correlations are also significantly stronger (e.g., digits are often used to enumerate text or other items rather than occurring with other digits). In a later study, we also trained the network by using an input set that satisfied these assumptions. In this case, the network self-organized to produce a segregated letter area, but individual digits were not grouped together. The

model thus predicts that spatially segregated cortical areas dedicated to letters should be more common and robust than areas dedicated to digits, as we observed in the functional neuroimaging experiment described.

BEHAVIORAL STUDY

If the co-occurrence hypothesis is correct, then people who have been exposed to letters and digits with different co-occurrence statistics might develop a different functional architecture. We explored this idea by using a behavioral measure known as the alphanumeric category effect. This effect refers to the fact that for most people, a letter "pops out" when presented in an array of digits compared with when it is presented in an array of letters, that is, it is detected faster and with less serial search (19–26). The cortical segregation of letter and digit recognition naturally accounts for this so-called alphanumeric category effect because the representations of letter distracters, which are in the same cortical region as the representation of the letter target, would interact with and potentially interfere with the target representation (via the short range cortical connections). Conversely, digit distracters, which are represented elsewhere, would presumably cause less interference. As a result, it would be easier to represent (and detect) a letter target when surrounded by digits than when surrounded by letters, and such a category effect is precisely what is found.

If letter segregation is caused by the co-occurrence of letters with other letters rather than with other types of stimuli, as the co-occurrence hypothesis assumes, then people who regularly process letters and digits together might show a reduced category effect. We tested this prediction experimentally by comparing the category effect in postal employees who process Canadian postal codes (in which letters and digits alternate, e.g., V5A 1S6) with postal employees who do not (27).

Fig. 4 shows the results. An alphanumeric category effect is evident for all subject groups as evidenced by the longer reaction times in detecting a letter among letters compared with a letter among digits. As predicted however, the Canadian mail sorters showed a smaller effect than postal worker controls who did not sort mail (measured both by the absolute difference in and ratio of response times in the letter-among-
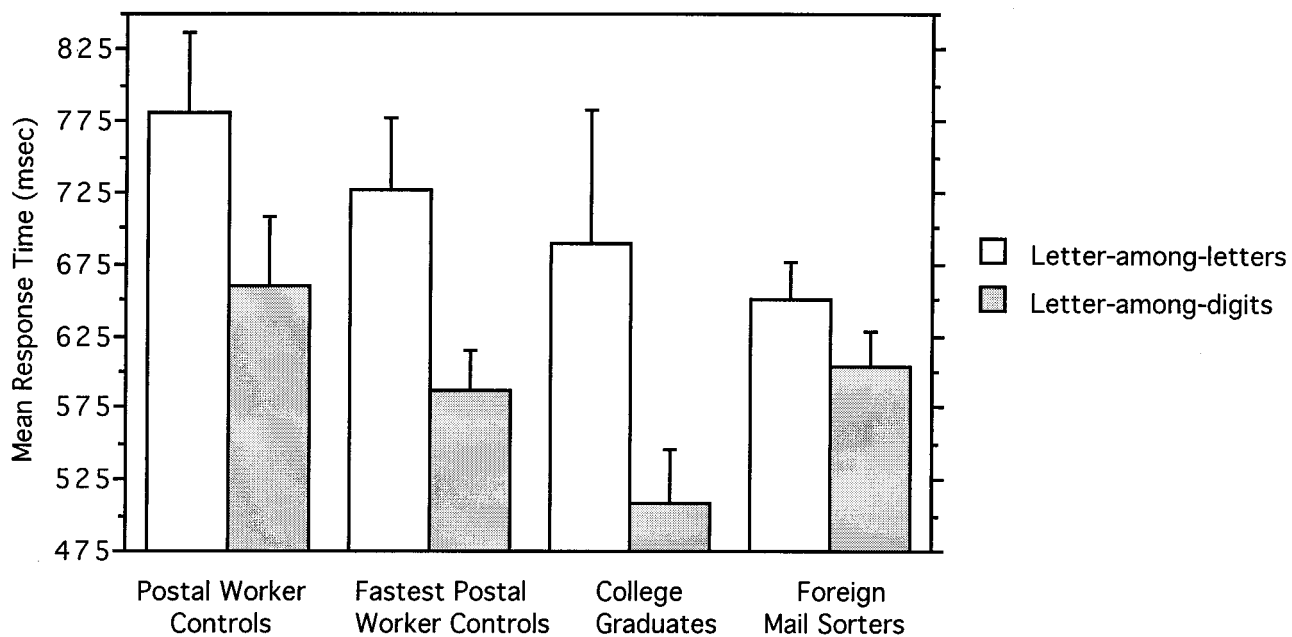

FIG. 4. Mean reaction time (with SE bars) for correct responses on target-present trials for 16 postal worker controls, the 13 fastest postal worker controls, 8 college graduates, and 10 foreign mail sorters.

letters and letter-among-digits conditions). The sorters were, however, faster than controls, presumably because of their extensive experience with speeded tasks. So, to ensure that these results were not the result of a floor effect, we excluded the three slowest postal worker controls (out of 16) for one analysis and used college graduates whose response times were faster for another. In both cases, the control group showed a larger category effect than the sorters even though both groups were faster than the sorters in the letter-among-digits condition. These cross-over interactions eliminate any obvious interpretations based on scaling artifacts and confirm the prediction of the co-occurrence hypothesis.

DISCUSSION

Letters and digits are distinguished, not by any obvious physical features, but solely by cultural conventions. Furthermore, the ability to recognize them is not innate. Nevertheless, we found evidence that the neural substrates underlying letter recognition are segregated from those underlying digit recognition in most normal subjects. The fact that letter and digit recognition depend on different neural substrates therefore suggests that the environment can lead to qualitative changes in the brain's functional organization. How might that happen? We have shown that a robust statistical property of the environment (the co-occurrence of letters), in conjunction with simple and widely accepted assumptions about the computational properties of cortex (correlation-based learning and lateral interactions), will lead to the segregation of such arbitrary and noninnate categories. We also have confirmed a critical prediction of this hypothesis, namely, that subjects exposed to a visual environment in which letters and digits occur together rather than separately would show less behavioral evidence of processing the two stimulus categories separately.

The co-occurrence hypothesis also may explain other counterintuitive examples of functional localization, such as the localization of musical processing and handwriting. Just as letters co-occur in text, musical sounds occur together in music and written characters occur together in writing. Indeed, even at the level of writing cursive vs. print, co-occurrence of stimuli is satisfied.

The point is not that the processing of any stimuli that co-occur in the environment will necessarily come to be localized in cortex. The hypothesis assumes both that the neural processing is local and that the relevant neural representations reflect the statistics of the environment (that is, the neural representations themselves co-occur). Complex stimuli with widely distributed representations presumably would not satisfy these constraints. Nevertheless, the co-occurrence hypothesis does offer a plausible new explanation for the localization of a number of arbitrary categories for which there is evidence of cortical specialization.

1. Alexander, M. P., Fischer, R. S. & Friedman, R. (1992) *Arch. Neurol.* **49,** 246–251.
2. Hanley, J. R. & Peters, S. (1996) *Cortex* **32,** 737–745.
3. Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G. & Belleville, S. (1994) *Brain* **117,** 1283–1301.
4. Gardner, H. (1974) *J. Psycholinguistic Res.* **3,** 133–149.
5. Allison, T., McCarthy, G., Nobre, A., Puce, A. & Belger, A. (1994) *Cereb. Cortex* **4,** 544–554.
6. Hebb, D. O. (1949) *The Organization of Behavior: A Neuropsychological Theory.* (Wiley, New York).
7. Polk, T. A. & Farah, M. J. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 12370–12373.
8. von der Malsburg, C. (1973) *Kybernetik* **14,** 85–100.
9. von der Malsburg, C. (1979) *Biol. Cybern.* **32,** 49–62.
10. Goodhill, G. J. (1992) Ph.D. thesis (University of Sussex, Brighton, U.K.).
11. Cottrell, M. & Fort, J. C. (1986) *Biol. Cybern.* **53,** 405–411.
12. Durbin, R. & Mitchison, G. (1990) *Nature (London)* **343,** 644–647.
13. Linsker, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7508–7512.
14. Linsker, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8390–8394.
15. Linsker, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8779–8783.
16. Miller, K. D., Keller, J. B. & Stryker, M. P. (1989) *Science* **245,** 605–615.
17. Ritter, H. (1990) *Psychol. Res.* **52,** 128–136.
18. Kohonen, T. (1988) *Self-Organization and Associative Memory* (Springer, New York), 2nd Ed.
19. Foldiak, P. (1991) *Neural Computat.* **3,** 194–200.
20. Duncan, J. (1980) *Psychol. Rev.* **87,** 272–300.
21. Duncan, J. (1983) *Percept. Psychophys.* **33,** 533–547.
22. Egeth, H., Jonides, J. & Wall, S. (1972) *Cognit. Psychol.* **3,** 674–698.
23. Jonides, J. & Gleitman, H. (1972) *Percept. Psychophys.* **12,** 457–460.
24. Merikle, P. M. (1980) *J. Exp. Psychol. Gen.* **109,** 279–295.
25. Schneider, W. & Shiffrin, R. M. (1977) *Psychol. Rev.* **84,** 1–66.
26. von Wright, J. M. (1972) *Scand. J. Psychol.* **13,** 159–171.
27. Polk, T. A. & Farah, M. J. (1995) *Nature (London)* **376,** 648–649.