

From M. L. Commons, R. J. Herrnstein, S. M. Kosslyn, & D. B. Mumford, (Eds.). 1990. *Quantitative Analyses of Behavior: Computational and Clinical Approaches to Pattern Recognition and Concept Formation*. Hillsdale, NJ: Erlbaum (Volume IX, Chapter 6, pages 109-126).

## A Quantitative Model for Pattern Recognition

*Eric G. Heinemann*

Brooklyn College of the City University of New York

*Sheila Chase*

Hunter College of the City University of New York

This chapter describes a model for visual pattern recognition by pigeons that is an extension of a much more general model for learning and behavior described in earlier chapters in this series (Chase, 1983; Heinemann, 1983a,b). The model we present belongs to the class of template-matching model. The comparison process that may be loosely described as “template matching,” however, this is only one of several basic processes with which this model deals in detail. These processes include the development of “focal attention,” and memory storage and retrieval processes. Note that most basic assumptions of the model were introduced to deal with phenomena quite far removed from pattern recognition. For example, the assumption that the decision process that leads to response selection is based on only a small sample of the information that is held in long-term memory plays an essential role in the account the model gives of the outcome of experiments on probability learning and matching (Heinemann, 1983a), and on the limit on the number of stimulus values that can be accurately identified (Chase, 1983).

### DESCRIPTION OF THE MODEL

The experiments to which our model was first applied were done with pigeons who were presented with one of two stimuli, S1 or S2, on each of a series of training trials. They were rewarded for making one response, R1, in the presence of one of these stimuli and for making an alternative response, R2, in the presence of the other. The learning that goes on in this situation is assumed to have two stages. The first stage is the presolution period (PSP). During the PSP the subject detects the statistical association that the experimenter (or nature) has arranged among the stimuli presented, the subject’s behavior, and the consequences of this behavior (e.g., reward or no reward). A quantitative model for the PSP has been presented by Heinemann (1983b) but is not considered in detail here. It is necessary to point out, however, that the processes occurring during the PSP have only an indirect effect on behavior. They lead the subject to discover and subsequently “attend to” relevant stimuli, i.e., features of the environment or other cues that predict outcomes of behavior. What we mean by the phrase “attend to” is that the subject processes information in the manner that characterizes the second stage of learning, as described in the paragraphs that follow.

It is assumed that, during the second stage of learning, important information gathered on each trial of an experiment is placed in a memory that has a large but not unlimited storage capacity. This memory is referred to as the long-term memory (LTM)<sup>1</sup>. Each record placed in

---

<sup>1</sup> We have previously referred to this as the limited capacity memory (LCM) in order to emphasize limited storage and overwriting. More recently (Chase, 2001) we have referred to this as exemplar memory to emphasize the notion that individual exemplars, instances or records are remembered.

this memory during the course of training contains information concerning the discriminative stimuli that the analysis done during the PSP has shown to be predictive of the outcomes of behavior, the “response” made, and the reward received. One such record is entered into the LTM on each trial and is said to occupy a “storage location.” The LTM is assumed to have a fixed number of storage locations, and the location to which each record is sent is selected randomly.<sup>2</sup> Any record occupying a storage location is destroyed (“overwritten”) when a new record is entered at that location.

The response made just before receipt of reward is assumed to be represented on the memory record in the form of sensory information associated with this response, such as the color of the key that was pecked, or the relation of the key to other features of the environment. To simplify matters, however, the responses are treated here as though they were represented in memory simply by the labels R1 and R2.

Information concerning reward and no reward is treated in a comparable fashion. The remembered sensory effects of rewards delivered repeatedly are assumed to be normally distributed. For situations in which the amount and kind of reward are not varied, however, it is sufficient to represent this information simply as the presence or absence of reward.

The most important way in which this model differs from our previous one is in the representation of the discriminative stimuli. Our previous theoretical work with visual stimuli dealt with their non-spatial properties, i.e., variations in luminance, dominant wavelength, and purity of patches of light. To deal with pattern recognition we assume that, to any distribution of light over the retina, there corresponds a continuous distribution of sensations, which we call the internal representation of the visual field. Instead of treating the distribution of sensations as continuous, however, we find it convenient to follow the computer-vision practice of considering the visual field partitioned into finite-sized cells called picture elements or *pixels*. Each pixel is assumed to have uniform brightness, hue, and saturation; thus each pixel may be treated in the same way that patches of light were treated in our earlier formulation.<sup>3</sup> Any particular visual experience can be represented in memory by the spatial coordinates of the relevant pixels, and codes for the brightness, hue, and saturation of each pixel. However, in this chapter we consider only patterns consisting of black pixels on a uniform white background. In this simplified situation we represent each pattern simply by the coordinates of the black pixels.

As mentioned previously, it is one of the assumptions of our model that the subject stores in LTM information concerning only those sensory events that the analysis done during the PSP has shown to be relevant. Our original treatment of the PSP did not consider the spatial aspects of visual stimuli, but it is necessary to consider this matter now to specify just what portions of the visual field are stored in LTM at the end of each trial. We shall assume that the statistical analysis the subject performs during the PSP (see Heinemann, 1983b) is done on the information represented at each location of the visual field separately and independently. Speaking loosely, each location (pixel) has its own PSP. Thus, only information from parts of the visual field that contain relevant sensations is represented on records stored in the LTM, and only such information is used in the decision processes described in the *Response Selection* section. Such selective processing of sensory information from relevant areas of the visual field has been referred to by others as “focal attention” (see Neisser, 1967).

---

<sup>2</sup> As noted by Heinemann (1983), “in future versions of this model it may be useful to assume that only a limited number of the... storage locations are available at any time, and that the loci of these storage locations drift slowly over time. Such an assumption might make it possible to deal with a variety of temporal effects such as spontaneous recovery and regression, trial spacing effects, etc. This general concept is similar to that incorporated in the theories of Estes (1955) and Landauer (1975)” (p. 19).

<sup>3</sup> The elementary areas of “pixels” of the internal representation do not interact. Interactions such as those evidenced in simultaneous contrast effects, formation of Mach bands, etc. are regarded tentatively, as the results of visual processes that help construct the internal representation of the visual field.

The positions of the relevant pixels should not be thought of as corresponding perfectly to the patterns presented for recognition. Rather, we assume there is some spread caused by factors such as variations in the location and size of the retinal image that result from trial-to-trial variations in head and eye position. An important consequence is that training to discriminate among several patterns presented successively to the same general region of the retina will result in focal attention to a region of the visual field that spreads out for some distance around the average position of the target patterns.

*Retrieval.* It is assumed that during each trial the subject retrieves a small random sample of positive records (records showing that a reward was received), from the LTM and that the only stored information that is used in the response selection process is the information contained in this sample of records.

*Response Selection.* The choice of response is based on a comparison of the information retrieved from LTM to the pattern of sensations that is present on a particular trial in those areas of the visual field that the analysis done during the PSP has shown to be predictive of outcomes. This pattern of sensations is called the *current input*. In this chapter we treat such current inputs and their representations in memory as static--as snapshots, freezing images in time and space. This is a simplification that we plan to treat as a special case when the model is developed further.

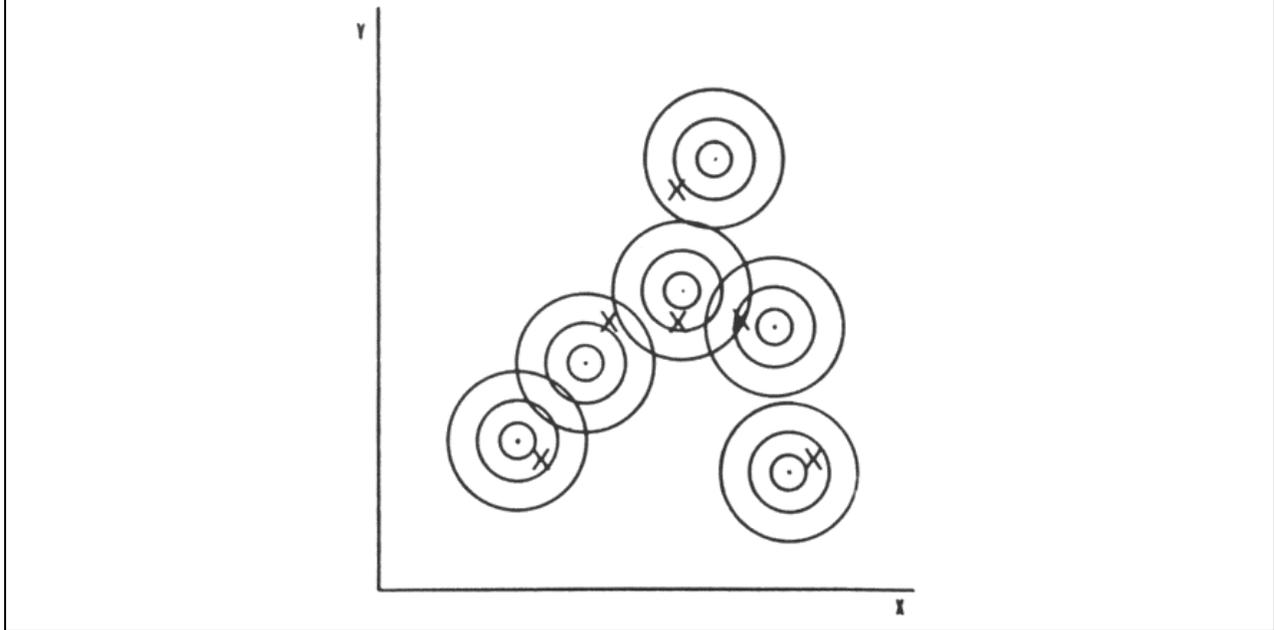
Before describing the assumed comparison process in detail, it is necessary to state two assumptions concerning memory variance: (1) while a record of a visual scene resides in LTM, Gaussian noise is added to the stored values of the vertical and horizontal coordinates of each pixel; (2) after a record has been retrieved (and is being held in working memory), the values of each pair of remembered coordinates fluctuate rapidly over time, momentary values falling into a bivariate truncated normal distribution with the means of the marginal distributions equal to the values retrieved from the LTM. Each of these distributions is truncated in the sense that probability densities at points that lie more than approximately five standard deviations to either side of the mean (of the marginal distributions) are treated as zero. This truncation has the effect of removing from consideration in the response selection process any records that represent severe mismatches.

We describe the assumed response selection process with reference to a specific example. A mathematical description of the process (for Gaussian distributions, ignoring the truncation) is presented in the Appendix.

Consider a situation in which a pigeon has been trained to identify each of  $n$  different dot-matrix letters by pecking on one of  $n$  response keys. On any particular trial the subject is shown a single letter for identification. According to our model, the subject first retrieves a small number of records from the LTM. The subject then compares the pattern it is now looking at with the pattern represented on each of the records retrieved from LTM to find the best match. Figure 6.1 illustrates one such comparison. The x's in Fig. 6.1 represent the current input (dot-matrix letter A), and the concentric circles represent the bivariate probability density functions corresponding to the points on the remembered letter. A measure of how similar the current input pattern is to the remembered pattern is calculated as follows.

*Step 1.* At each point in space at which a current input pixel is located, calculate the mean of the probability densities that represent each of the points on the memory record. This process will yield a number of means equal to the number of input points. Next multiply these means together, the resulting product to be called JD (joint density).

**Figure 6.1.** Remembered group of points representing the letter A as shown on a single record retrieved from LTM. The concentric circles represent contours of constant probability density on the bivariate distributions for the spatial coordinates of each point. The x's represent points on the current input.



*Step 2.* Repeat this process for each of the records retrieved from memory. Add the JD values of all records that carry the same response label. The decision rule is: Make the response that is associated with the largest value of JD. If none of the JD values is greater than zero, draw a new sample of records from the LTM. Finally, if repeated sampling fails to yield useful information (i.e., a JD value greater than zero), simply make the response associated with the greatest probability of reward in the past.

The assumption that the distributions that represent pixels of remembered scenes in working memory are truncated insures against the acceptance of one kind of really bad mismatch: If even a single point of the current input is so far away from the points represented on a particular memory record that the mean probability density at that input point is equal to zero, then the record under consideration makes no contribution to the selection of the response with which it is associated because its JD value is equal to zero. An analogous kind of mismatch could occur if the mean probability density at all the current input points were greater than zero but the memory record contained an outlying point. This case is treated in essentially the same way as the first, namely, the memory record is treated as though JD were equal to zero.

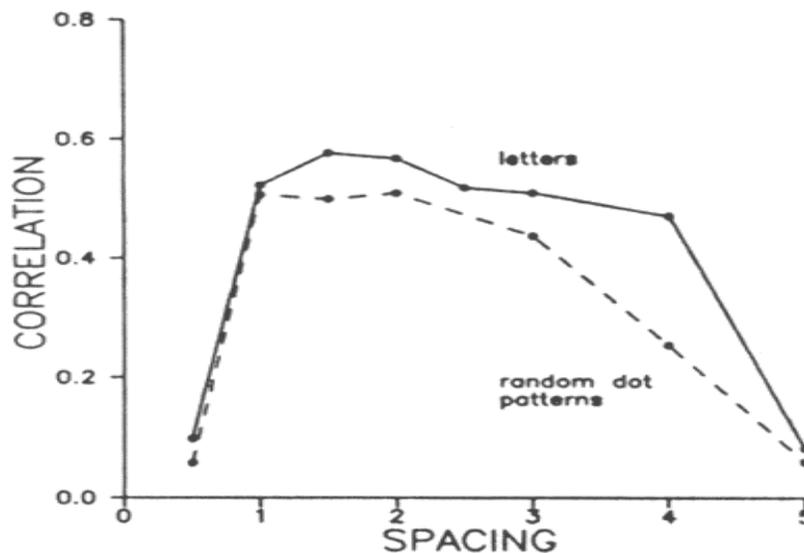
We have represented this model in a computer program that we use to simulate the behavior of pigeons in a variety of situations.

## SELECTED APPLICATIONS OF THE MODEL

### Pattern Confusion Matrices

Several experimenters have shown that pigeons can be trained to recognize letters in various type-fonts as well as dot-matrix letters on computer screens (D.S. Blough 1982, 1985; Lea & Ryan, 1983; Morgan, Fitch, Holman, & Lea, 1976). In a very comprehensive study, D.S. Blough (1985) obtained a confusion matrix for all 26 letters of the alphabet. To test our model against these data, we used our simulation program to generate a number of letter confusion matrices. The only parameter varied in these simulations was a multiplier of the spatial coordinates that we refer to as “spacing.” This parameter determines the distance between the input points (the x’s in Fig. 6.1) and between the peaks of the distributions that represent the remembered points. Figure 6.2 shows the product-moment correlation between the computer-generated and pigeon-generated matrices as a function of spacing (solid line). The best correlation is about 0.58, which is virtually identical to the correlation Blough found with a matrix predicted by a fuzzy template model he developed.

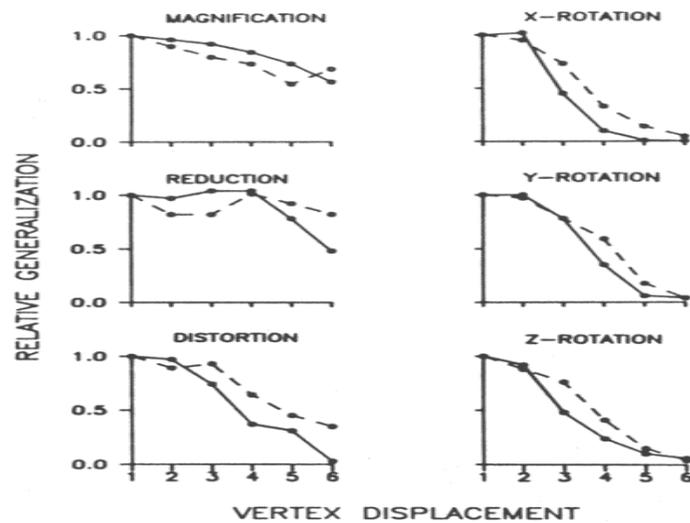
**Figure 6.2.** The correlation ( $r$ ) between the empirical and theoretical confusion matrices, as a function of the spacing parameter of the simulation program (see text). The solid line refers to confusions of letters, the dashed line to confusions of random dot patterns. The empirical data are from D.S. Blough (1985).



D.S. Blough (1985) did a similar experiment using 16 different random dot patterns in place of letters. The dashed line shown in Fig. 6.2 shows the correlation between the confusion matrices generated for this situation by the pigeons and by our simulation program. The maximum correlation obtained differs only slightly from that obtained for letters. This is an important finding because the model we are presenting must be able to deal with confusions among any set of arbitrary shapes. To the extent that it does so, it has a clear advantage over the principal theoretical alternative, namely, models based on feature analysis. Proponents of feature analysis appear not to have agreed on a set of features underlying letter recognition, much less the recognition of arbitrary shapes.

Letter-confusion data are often analyzed by computer programs such as CLUSTER (SAS Institute, 1982). The output of such programs can guide and constrain guesses concerning features that may underlie the confusions, but this has little if any bearing on questions concerning the existence of feature detectors or the role of a possible feature extraction process in recognition. As D.S. Blough (1985) points out: “it would be a mistake to conclude that feature analysis is actually involved in the discriminative processes underlying the data. Fits of feature models depend strongly on the features chosen for input, but the choice of features depends, directly or indirectly, on data such as those being explained” (p. 270).

**Figure 6.3.** Relative generalization to various transformations of a two-dimensional shape measured by Cerella (1990). The solid lines represent the mean results for Cerella’s pigeons, the dashed lines the results of simulations. The horizontal axes of the various panels have a common scale, a measure of prototype distortion in terms of vertex displacement. See Cerella for details.



## Generalization

The data discussed in the preceding paragraphs represent confusions between patterns observed after the subject has had prolonged training to discriminate among the patterns. We consider next some results obtained in generalization experiments in which subjects are typically trained to discriminate between two patterns and are then tested, without informative feedback, for recognition of one of the patterns in altered sizes, positions, angular orientations, etc.

Data presented in Volume 8 of this series (Commons, Herrnstein, Kosslyn & Mumford, 1990) suggest that changes in the size and position of two-dimensional forms are less disruptive of recognition performance than rotations or distortions of these forms (see Cerella, 1990; Lombardi & Delius, 1990) or the addition of distractors (see Steele, 1990). For example, Cerella studied relative generalization of responding to a simple form (prototype) when that form was shifted to a new position on the display surface, reduced or magnified in size, distorted

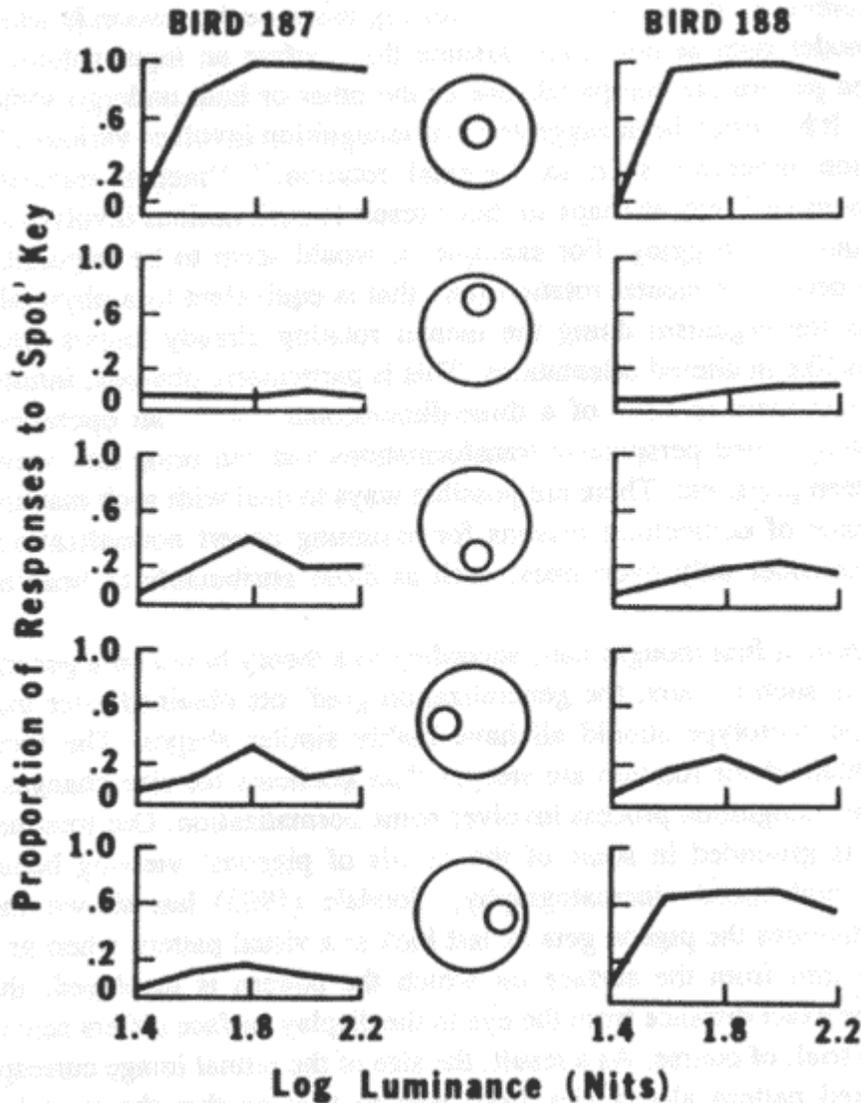
randomly, or distorted as it would be if it were the projection on a plane of a solid figure rotated about its x-, y-, z-axis (see Cerella for further details). His data, reproduced in Fig. 6.3, show that the gradients for reduction and magnification have more extended horizontal segments than do the others.

In the context of these results an interesting theoretical question is whether a template model such as ours must assume that, before an input pattern and a remembered pattern are compared, one or the other or both undergo some normalization. It has often been suggested that recognition involves various internal normalization processes such as “mental rotation,” “mental translation,” “mental zooming,” etc. Perhaps so, but a resort to such notions involves a great danger of question begging. For example, it would seem to be impossible, in general, to perform a mental rotation (one that is equivalent to a physical rotation) unless the organism doing the mental rotating already knows what the object looks like in altered orientations. This is particularly obvious, intuitively, when one considers rotation of a three-dimensional object, an operation that produces complicated perspective transformations and can bring into view previously unseen parts, etc. There are possible ways to deal with such matters, but in the absence of compelling reasons for assuming covert normalization processes we consider only overt ones, such as those attributable to head movements.

It may seem at first thought that, according to a theory based on a generalization principle such as ours, the generalization gradients obtained after training with a single prototype should all have highly similar shapes. The fact that gradients obtained for rotation are steeper than gradients for size changes suggests that the recognition process involves some normalization. Our treatment of this matter is grounded in some of the details of pigeons’ viewing behavior. Through high-speed cinematography, Goodale (1983) has shown that in go/no-go situations the pigeon gets its last look at a visual pattern when its head is about 55 mm from the surface on which the pattern is displayed, the F2 position. The exact distance from the eye to the display surface differs somewhat from trial to trial, of course. As a result, the size of the retinal image corresponding to a fixed pattern also varies from trial to trial so that the records that represent a given stimulus pattern in memory represent a distribution of sizes rather than a single size. On the basis of the measures of the variance of F2 provided by Goodale, it is possible to compute the variance of the trial-to-trial distribution of retinal image sizes. When this is taken into account, our model yields the fits to Cerella’s data shown in Fig. 6.3. The solid lines represent Cerella’s data, and the dashed lines represent the outcome of simulations. The results of the simulations do not represent “best fits” but crude approximation obtained by trial and error. Only the “spacing” parameter was varied in this process.

Variations in head position do not affect the form of the retinal image produced by a two-dimensional representation of an object in the same way they would affect the form of the retinal image produced by the object itself. For this reason training with two-dimensional representations of three-dimensional forms will not produce the kind of remembered form distributions that can mimic normalization. It is not surprising, therefore, that the gradients for figures “rotated” around the x, y, or z axes (Fig. 6.3) drop off much more rapidly than gradients for figures varying in size.

**Figure 6.4.** Choice curves obtained when the spot was in the positions shown. The spot and the surrounding field are not drawn to scale; the diameter of the spot was .5 cm; that of the surrounding field was 7.62 cm. The pigeons were trained to choose one key in the presence of the spot and the other key in its absence. The luminance of the spot was 1.8 log Nits and the luminance of the large field was 1.4 log Nits. From Heinemann & Kadison (1976).



(1976)

The effects of varying the location of the prototype on the display surface depend on the conditions of training. If the pattern is moved about from trial to trial during training, pigeons apparently learn to search for it and peck near the target pattern wherever it is (D.S. Blough, 1979; P.M. Blough, 1984; Jenkins & Sainsbury, 1969). However, if the patterns to be discriminated appear always in the same location with respect to the visual framework, say in the center of the key, then pigeons behave as though they fixate a central region defined by the visual frame. Figure 6.4 shows the results of a generalization test given to pigeons who had been trained to peck on one disk when presented with a small circular patch of light projected on a larger evenly illuminated field, and to peck on a different disk when presented with the large field without the added spot (Heinemann & Kadison, 1976). The visual display was located between the two choice keys. The luminance and position of the added spot were varied during

the generalization test. Notice (Fig. 6.4) that the discrimination between the presence and absence of the spot was essentially perfect when the spot luminance was reasonably high and the spot was in the center of the field. For most of the other test positions, however, the bird behaved almost as though the spot did not exist.

It might be objected that this experiment is not relevant because it dealt with brightness recognition rather than pattern recognition. However, we have obtained essentially similar results in several unpublished experiments on pattern recognition. In one of these, three pigeons were trained to identify each of four visual shapes by pecking on one of four different response keys. During training the single pattern shown on each trial was always presented near the center of the 6 cm square face of a small monochrome TV monitor. During a “translation test” that followed 130 days of training (100 trials per day), “probe” trials were interspersed among the training trials. On each probe trial one of the four shapes was presented 3.5 cm above and to the right of the training position. During the test the birds averaged 72% correct when the stimuli were in their training position but only 31% correct when the stimuli were in the new positions; two of the birds actually performed at chance (25% correct) when the stimuli were shown in the new positions. The position change studied here was quite large ( $32^\circ$  of visual angle assuming the average distance between the pattern and the pigeons head to be 55 mm). Cerella (1990) studied the effect of much smaller changes and found that there is a small range of translations over which recognition performance does not change much. Our model accounts for these location effects as follows: When the location of the target pattern is fixed during training, focal attention should be confined to the general neighborhood of the target pattern. Within this region the pattern may be moved without effect. However, when the pattern the subject has been trained to recognize is translated over a substantial distance, more specifically, if it has been moved out of the region of focal attention, our model predicts that the subject will fail to recognize it. As noted before, this somewhat counterintuitive prediction has been confirmed in several experiments.

### Distractor Effects

Under many conditions the ability to recognize a particular pattern deteriorates as a result of introducing additional noninformative patterns in the visual field.

What is meant by noninformative is that the identical pattern, such as a dash, is added to each of the patterns that the subject is expected to recognize. “Distractor” effects of this kind are discussed by Steele (1990). In one spectacular demonstration, Steele shows that letters that are recognized quite well when shown alone on a uniform background are no longer recognized at all when the background is filled with dashes, and it appears that the recognition task has to be relearned.

The way our model accounts for these phenomena is as follows: First, dashes that lie outside the region subject to focal attention will have no effect on recognition performance. Those dashes that lie close to the letters will be processed as part of the current input and will result in a distractor effect that involves two mechanisms. On the first trial on which dashes are added to the letters that the subject has learned to recognize, the subject’s memory contains no records of patterns that include dashes, and the comparison process described previously will yield joint densities equal to zero. This will result in resampling and eventual guessing, i.e., selecting a response on the basis of remembered reward probabilities alone. As retraining proceeds under the new conditions, the memory will come to contain more and more records that represent letters plus dashes, and guessing will soon cease to be a significant factor. According to our model, however, distractor effects will still be observed in subjects who have been fully trained with patterns that include a distractor. The reason for this is that the presence of the distractor increases the variance of the quantities that determine response selection, without contributing to

discrimination. Computer simulations have shown that the expected distractor effect decreases with increasing distance between the target and distractor. Such a distance effect was in fact observed in the experiments of Steele (1990).

Some observations made in our own laboratory support and supplement those of Steele (1990). After completing the translation test described in a previous paragraph, we did a distractor test with the same pigeons. A period of retraining intervened between these two tests. During the distractor test a small circle, roughly the same size as the training stimuli, was presented near each of the training patterns. When the circle was close to the training patterns (3 mm from the nearest border), the birds averaged 58% correct in this four choice absolute identification task. When the distractor was absent or was 3.5 cm distant from the training pattern, the birds averaged 74% correct. The earlier translation test had shown that if the training patterns were shifted to the same location, 3.5 cm from the training one, the training patterns were no longer recognized, or poorly recognized, depending on the bird. The notion of focal attention implies that locations in which translated training patterns are no longer recognized should also be the locations in which distractors no longer distract. The results mentioned are compatible with this implication but more detailed research needs to be done on this matter.

## DISCUSSION

Template models such as that presented here have not been popular among psychologists, who have generally preferred some form of feature extraction approach. By and large, however, feature theories have not had much success as measured in terms of predictive power. As noted earlier, this is partly because it has not been possible to find a set of features that can predict results for any arbitrary pattern. Further, as pointed out by Steele (1990), feature theories encounter extreme difficulties when attempting to deal with context effects, e.g., distractor effects. It seems that the features of a given pattern may change as the visual context changes; that there are features that emerge in a given context, etc. To be useful, then, feature theories must provide rules that specify how features combine or interact or change in various contexts. It may well be possible to do this, but someone familiar with the history of Gestalt Psychology might not be overly optimistic.

Why then are template models unpopular? Among the factors that appear to be responsible are a widespread failure to appreciate the vast capacity of the pigeons' LTM and the notion that some type of top-down processing is required. The model we have presented requires that the pigeons have a relatively large memory for visual scenes. Recent evidence shows that they do indeed. Vaughan and Greene (1983) trained pigeons to memorize, by rote, more than 300 scenes projected in full color. In a related unpublished experiment by Heinemann, Stevens, Ionescu, and Neiderbach, full-color slides of natural scenes were projected on a small screen located between two choice keys. Pigeons were rewarded for a sequence consisting of several pecks at the projected picture followed by a peck on either the left or right choice key. Half of all pictures shown were randomly assigned for reinforcement on the right, the other half of reinforcement on the left. During the course of the experiment the number of pictures presented was increased in rather large steps up to 640, at which point the pigeons were performing at a level of about 85% correct. Thus, no doubt exists that pigeons can memorize a surprising number of visual scenes, certainly a number that is compatible with the number of storage locations assumed in most of our simulations (cf. Chase, 1983; Heinemann, 1983a).

Some of the preference for feature over template models springs from the notion that there is convincing evidence that demands explanation in terms of top-down processing, e.g., the suggestion that recognition involves normalization operations such as mental rotation, mental zooming, etc. It may be a very useful strategy to try to account for the phenomena observed in

pattern-recognition situations in terms of somewhat simpler, well-understood processes. For example, as we have shown, a template model can account for mental zooming simply by taking into account observed variations in the distance between the pattern and the position of the head at the time a record is placed into LTM.

Data suggesting that pigeons have concepts can also be easily accounted for by a model such as ours. In concept experiments with pigeons, the birds are usually presented with a relatively large number of different instances and non-instances of the concept and then tested with new exemplars of each. They are said to “have” the concept if they classify the new instances appropriately. This is exactly what would happen if, when novel exemplars are presented, the pigeon samples its memory until a reasonable approximation to a match is found. One would expect better transfer of the concept to new instances as the number of exemplars in LTM increases because this would increase the probability of finding a reasonable match. There is some evidence indicating that this is indeed what happens (Greene, 1983).

The conceptual conflict between a model of the type we have presented and more cognitive models has its counterpart in the field of artificial intelligence. Traditionally, attempts to simulate human intelligence, for example, to have computers play checkers or do medical diagnoses, have relied on expert programs. The procedure is to learn as much as possible about the rules followed by human experts and then write a program that applies these rules. There is general agreement that expert programs in such fields as checkers, chess, and medical diagnosis are far from achieving the level of performance of human experts. At least some writers on artificial intelligence have argued recently that the approach is quite hopeless (Dreyfus & Dreyfus, 1984). Their point is that rules are for beginners, e.g., novice chess players or medical students. What distinguishes real human experts is that they rely on their memory for an incredibly large number of special cases.

## ACKNOWLEDGMENTS

The research reported here was supported by NIMH Grant Number 5 RO1 MH40712, PSCUNY Grants Number 6-65358 and 6-66189, and by computing resources provided by the City University of New York, University Computer Center.

## REFERENCES

- Blough, D.S. (1979). Effects of the number and form of stimuli on visual search in the pigeon. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 211-223.
- Blough, D.S. (1982). Pigeon recognition of letters of the alphabet. *Science*, 218, 297-298.
- Blough, D.S. (1985). Discrimination of letters and random dot patterns by pigeons and humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 261-280.
- Blough, P.M. (1984). Visual search in pigeons: Effects of memory set size and display variables. *Perception & Psychophysics*, 35, 344-352.
- Cerella, J. (1990). Shape constancy in the pigeon: The perspective transformations decomposed. In M. L. Commons, R. J. Herrnstein, S. M. Kosslyn, & D. B. Mumford (Eds.). *Quantitative analyses of behavior: Behavioral approaches to pattern recognition and concept formation* (Vol. VIII: pp. 145-164). Hillsdale, NJ: Erlbaum.
- Chase, S. (1983). Pigeons and the magical number seven. In M.L. Commons, R.J., Herrnstein, & A.R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (Vol. IV: pp. 37-57). Cambridge, MA: Ballinger.

Dreyfus, H. & Dreyfus, S. (1984). Mindless machines: Computers don't think like experts and never will. *The Sciences*, Nov./ Dec., 18-22.

Estes, W.K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, 62, 145-154.

Greene, S.L. (1983). Feature memorization in pigeon concept formation. In M.L. Commons, R.J. Herrnstein, & A.R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (Vol. IV: pp. 209-229). Cambridge, MA: Ballinger.

Goodale, M.A. (1983). Visually guided pecking in the pigeon (*Columba livia*). *Brain, Behavior and Evolution*, 22, 22-41.

Heinemann, E.G. (1983). A memory model for decision processes in pigeons. In M.L. Commons, R.J. Herrnstein, & A.R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (Vol. IV: pp. 3-19). Cambridge, MA: Ballinger.

Heinemann, E.G. (1983b). The presolution period and the detection of statistical associations. In M.L. Commons, R.J. Herrnstein, & A.R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (Vol. IV: pp. 21-35). Cambridge, MA: Ballinger.

Heinemann, E.G. & Kadison, K. (1976). Control of pigeon's choice behavior by the position and luminance of a spot of light. *Bulletin of the Psychonomic Society*, 7, 522-524.

Jenkins, H. M., & Sainsbury, R.S. (1969). The development of stimulus control through differential reinforcement. In N. J. Mackintosh & W.K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 123-161). Halifax: Dalhousie University.

Landauer, T.K. (1975). Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology*, 7, 495-531.

Lea, S.E.G. & Ryan, C.M.E. (1983). Feature analysis of pigeon's acquisition of concept discrimination. In M.L. Commons, R.J. Herrnstein, & A.R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (Vol. IV: pp. 239-253). Cambridge, MA: Ballinger.

Lombardi, C. M. & Delius, J. D. (1990). Size invariance in visual pattern recognition by pigeons. In M. L. Commons, R. J. Herrnstein, S. M. Kosslyn, & D. B. Mumford (Eds.). *Quantitative analyses of behavior: Behavioral approaches to pattern recognition and concept formation* (Vol. VIII: pp. 41-66). Hillsdale, NJ: Erlbaum.

Morgan, M.J., Filch, M. D., Holman, J.G., & Lea, S.E.G. (1976). Pigeons learn the concept of an "A." *Perception*, 5, 57-66.

Neisser, U. (1967). *Cognitive psychology*. NY: Appleton-Century-Crofts.

SAS Institute. (1982). SAS system (Computer program). Cary, NC: Author.

Steele, K. M. (1990). Configural processes in pigeon perception. In M. L. Commons, R. J. Herrnstein, S. M. Kosslyn, & D. B. Mumford (Eds.). *Quantitative analyses of behavior: Behavioral approaches to pattern recognition and concept formation* (Vol. VIII: pp. 111-128). Hillsdale, NJ: Erlbaum.

Vaughan, W. Jr. & Greene, S.L. (1983). . In M.L. Commons, R.J. Herrnstein, & A.R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (Vol. IV: pp. 231-237). Cambridge, MA: Ballinger.

This article is reproduced from the book with some minor changes, e.g. in formatting and in references to other chapters in the series. The research reported here was supported by NIMH Grant Number 5 RO1 MH140712, PSC-CUNY Grants Numbers 6-65358, and by computing resources provided by the City University of New York, University Computer Center.

## APPENDIX

### Response Selection

*A. One-dimensional Stimulus Continua.* Consider first the case in which the sensations to be classified vary along a single dimension,  $x$ , when  $x$  might represent the brightness of a light, for example. Only one specific value of the variable  $x$ , denoted  $X$ , is presented on any trial of the experiment.

The task is to decide which response,  $R_c$ , ( $c = 1, 2, \dots, C$ ), is most likely to be correct when the subject is experiencing a particular sensation value,  $X$ . In this context  $X$  is referred to as the “current input.”

The LTM contains a large number of “records,” each of which contains information about events that occurred on a single trial of the experiment. Among other things, each record contains a description of the response that was made, referred to here as the “response label,” and a representation of the sensation induced by the stimulus that was presented on the trial. To be more specific about the latter, the memory of a particular current input,  $X$ , is represented in LTM by a random variable,  $h$ , which has a Gaussian distribution and a mean equal to  $X$ . (The stored value of the current input may be thought of as fluctuating over time, the Gaussian distribution representing momentary values.)

The process of response selection begins by copying a random sample of  $M$  records from the LTM. The remembered sensation value shown on each of these records, denoted  $H_i$  ( $i = 1, 2, \dots, M$ ), is the value of the random variable  $h$  at the moment of retrieval, i.e., each  $H_i$  represents a value of  $h$  randomly selected from the distribution represented on the LTM record.

After retrieval, Gaussian noise  $n$  is added to each constant  $H_i$ , so the remembered sensations are represented as random variables:

$$\begin{aligned} y_1 &= H_1 + n \\ y_2 &= H_2 + n \\ &\vdots \\ y_i &= H_i + n \\ &\vdots \\ y_M &= H_M + n \end{aligned}$$

Here  $n$  represents independent Gaussian noise added to each constant  $H_i$ . This representation of retrieved records is illustrated in Fig. 6.5. The response selected is the one associated with  $H_i$  in the following expression:

$$i = \max_{1, 2, \dots, M} \left[ p(y_i = X | H_i) = \frac{e^{-\frac{(X - H_i)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \right] \quad (1)$$

In the case in which some of the  $H_i$  are associated with the same response labels there are just  $C$  different labels, where  $C < M$ . Let  $H_{c1}, H_{c2}, \dots, H_{cs}, \dots, H_{cSc}$  denote the  $S_c$  different  $H_i$ 's that

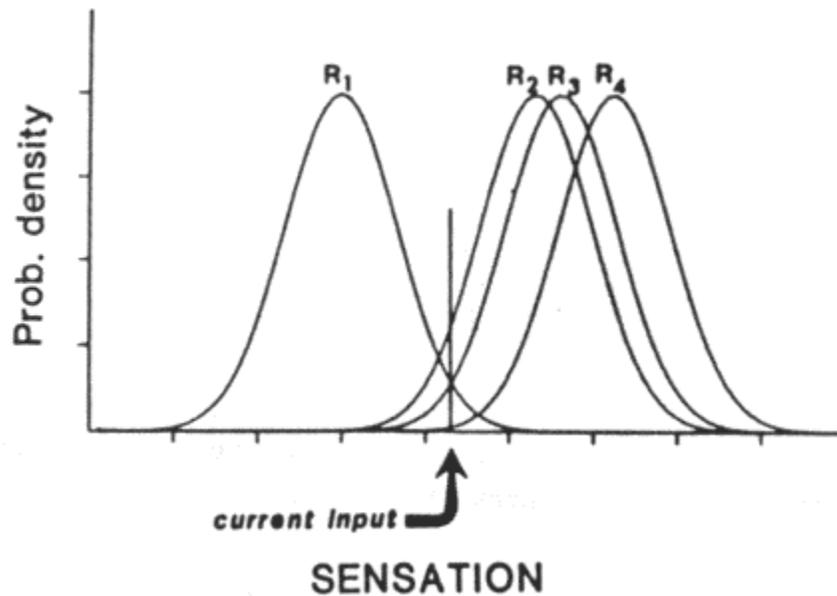
have the same response label  $c$ . In this case the probability densities at  $X$  are summed over the number of records,  $S_c$ , that have the same response label, before the maximum is obtained. Specifically, the decision criterion for this situation is:

$$c = \max_{1,2 \dots C} \left[ \sum_{s=1}^{S_c} p(y_{cs} = X | H_{cs}) = \sum_{s=1}^{S_c} \frac{e^{-\frac{(X - H_{cs})^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \right] \quad (2)$$

*B. Visual Patterns.* Visual patterns are represented by a sample of evenly spaced points. A particular current input,  $X$  (pattern to be identified), consists of  $K$  points:  $X = x_1, x_2 \dots x_k, \dots x_K$ .

Each record of a pattern retrieved from LTM,  $H_{ij}$ , is characterized by  $J_i$  points. The subscript  $ij$  refers to the  $j$  point on record  $H_i$ :  $H_i = H_{i1}, H_{i2} \dots, H_{ij}, \dots, H_{iJ_i}$ .

**Figure 6.5.** A sample of four records each associated with a different response,  $R_c$ . The response chosen is the one for which the probability density at the current input,  $X$ , is highest



Each point on the current and remembered patterns is characterized by its spatial coordinates,  $u$  and  $v$ . In addition, each point may be characterized by its value along the dimensions of hue, saturation, and brightness. In such multidimensional situations the current inputs and the sensation values shown on retrieved records are treated as vectors,  $\mathbf{X}$  and  $\mathbf{H}_i$ . The vectors all have the same number of coordinates, which is assumed to be fixed at  $N$ .

$$\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots, \mathbf{x}_k$$

$$\text{in which } \mathbf{x}_k = \mathbf{x}_{k1}, \mathbf{x}_{k2}, \dots, \mathbf{x}_{kn}, \dots, \mathbf{x}_{kN}$$

and

$$\mathbf{H}_i = \mathbf{H}_{i1}, \mathbf{H}_{i2}, \dots, \mathbf{H}_{ij}, \dots, \mathbf{H}_{iJ_i}$$

in which  $\mathbf{H}_{ij} = H_{ij1}, H_{ij2}, \dots, H_{ijn}, \dots, H_{ijN}$

As was the case for one-dimensional stimulus continua, the M records of sensations in the sample are represented by random variables:

$$\mathbf{y}_1 = \mathbf{H}_1 + \mathbf{n}$$

$$\mathbf{y}_2 = \mathbf{H}_2 + \mathbf{n}$$

⋮

$$\mathbf{y}_i = \mathbf{H}_i + \mathbf{n}$$

⋮

$$\mathbf{y}_M = \mathbf{H}_M + \mathbf{n}$$

Note that each value of  $y_i$  may be characterized by a different number of points,  $J_i$ , depending on the record,  $\mathbf{H}_i$ . In choosing a decision criterion we must deal with the problem that K, the number of points in the current input,  $\mathbf{X}$ , is not necessarily equal to  $J_i$ , for  $i = 1, 2, \dots, M$ . The rule used in our model is based on the number of points in X as shown in the following:

$$i = \max_{1,2,\dots,M} \left[ \prod_{k=1}^K \left\{ \sum_{j=1}^{J_i} \frac{1}{J_i} \left( P(\mathbf{Y}_{ij} = \mathbf{X}_k | \mathbf{H}_{ij}) \right) \right\} \right] \quad (3)$$

The probability density at each point designated by subscript  $k$  of  $X_k$  is averaged over all points  $J_i$ , for each record. This probability density function has N equally distributed Gaussian components of variance,  $\sigma^2$ , and is thus given by:

$$P(\mathbf{Y}_{ij} = \mathbf{X}_k | \mathbf{H}_{ij}) = \frac{e^{-\sum_{n=1}^N \frac{(X_{kn} - H_{ijn})^2}{2\sigma^2}}}{\left(\sqrt{2\pi\sigma^2}\right)^N} \quad (4)$$

Finally, for the case in which more than one record has the same response label,  $\max I$ , in (3) is replaced by  $\max c$ , and the term between the outmost brackets is summed over the number of records that have the same response label, yielding an expression analogous to (2).