Title	"Concept formation" and categorization by pigeons
Sub Title	
Author	Chase, Sheila
Publisher	慶應義塾大学大学院社会学研究科
Publication year	1997
Jtitle	慶応義塾大学大学院社会学研究科紀要：社会学心理学教育学 (Studies in sociology, psychology and education). No.45 (1997.) ,p.11- 17
JaLC DOI	
Abstract	
Notes	シンポジューム：Pattern recognition in humans and animals
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN0006957X-00000045-0011

"CONCEPT FORMATION" AND CATEGORIZATION BY PIGEONS

Sheila Chase*

Statement of the Problem

A large number of experiments done during the past half-century were intended to deal with the question of whether non-human animals form "concepts." Most of these experiments were done with pigeons but some have used monkeys and apes. In typical experiments on concept formation the animals are presented with examples of instances and non-instances of the concept (e.g. photographs including people or not including people) and are trained to make one response to instances and an alternative response to non-instances. In a subsequent test they are shown new exemplars of instances and non-instances. These new instances are categorized at greater than chance accuracy. The evidence is not controversial. Non-humans perform such tasks successfully (e.g., Herrnstein; 1979, D' Amato & Van Sant, 1988). What is needed now is an understanding of how such behavior comes about.

In further consideration of his matter it is important to keep in mind that the terms "instances" and "non-instances" used in the previous paragraph refer to entire photographs. Though the pigeons' successful performance in the categorization task is not in question, the assumption that the pigeons' responses are necessarily controlled by the rule the experimenter used in categorizing the stimuli is.

There is evidence that categorization of new exemplars may be based on details of the photographs other than those that governed the experimenter's category assignments. For example, Greene (1983) has shown that conceptually irrelevant aspects of the photographs, such as the background, may have control over the (pigeons') behavior. To give another example: D'Amato and Van Sant (1988) analyzed misclassifications in an experiment with monkeys in which good transfer to new exemplars was

_____
* Hunter College of the City University of New York, New York

found. Their analysis "raised serious questions about the degree to which their transfer behavior was governed by conceptual processes." (p. 43). For example, they found that photographs which to humans clearly do not show a person, e.g. a marine creature with two red spots on its body, were consistently classified as persons. In many of the photographs showing humans there were red patches on clothing, suggesting to D'Amato and Van Sant that exposure to these photographs was the source of misclassifications of non-person photographs having red patches.

It may well be that in many experiments the animals' categorization of stimuli is indeed based on the category specified by the experimenter, but that is usually very difficult to demonstrate. I will argue that the results discussed above can be accounted for by the notion that the subjects in these experiments memorize a large number of exemplars and classify new exemplars on the basis of their similarity to those in memory. In offering this account I do not mean to exclude the possibility that organisms may be genetically predisposed to categorize ecologically significant events or objects. It is possible that a number of different processes affect categorization. The task now is to understand these processes. In this paper I will show how our model, the Heinemann-Chase model of memory and decision processes, deals with categorization of complex stimuli by treating concept formation as a problem in pattern recognition. While this model has not yet been tested on the type of pictorial stimuli used in most concept formation experiments, it has been shown to describe recognition of outline figures by pigeons in precise quantitative detail. We are currently extending our model to treat early vision (Heinemann & Chase, 1995), the processes through which the physical stimulus is transformed into sensations. We intend to use this model in the analysis of the pictures used in
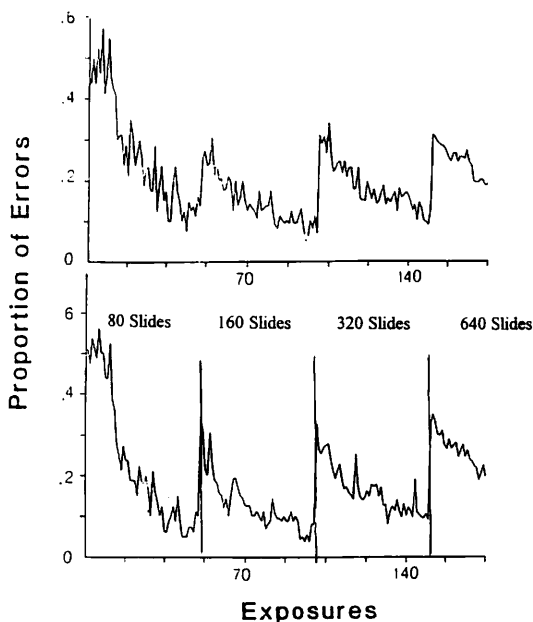
Fig. 1. Proportion of errors as a function of training. The stimuli were colored photographs. The response correct in the presence of each slide remained constant. There was no unifying concept. Half of the photographs assigned as correct for each key showed one or more persons, the other half did not. Eighty slides were presented during a session. Exposures refers to number of times each slide was seen. The data are shown for the two birds trained under this condition.

concept formation experiments.

The ability of pigeons to categorize an amazingly large number of colored photographs on the basis of rote memory is well documented (Vaughan & Greene, 1983; Vaughan & Greene, 1984). In a related experiment using choice rather than rate of response as dependent variable, Heinemann, Ionescu, Stevens, and Neiderbach (unpublished) showed pigeons photographs of natural scenes projected on a small screen located between two choice keys. The pigeons were rewarded for pecking either the left or right choice key. Half of all pictures shown were randomly assigned for reward if the right key was pecked, the other half for reward if the left key was pecked. During the

course of the experiment the number of pictures presented was increased in large steps from 80 to 640, at which point the pigeons were performing at a level of about 85 percent correct. Although the experiment was terminated at that point, there was no evidence that the pigeons were approaching a limit to the number of photographs they could categorize. The proportion of errors made by these birds somewhat before training was terminated are shown in Fig. 1.

This experiment, and those of Greene (1983) and Greene and Vaughan (1983; 1984), among others, show that an underlying rule is not necessary for categorization of colored photographs. However, when the photographs are unified by a rule (concept) that is meaningful to humans, categorization may be somewhat facilitated (Herrnstein, 1979).

According to our model, this facilitatory effect is a consequence of the fact that in a concept-learning experiment the stimuli categorized as "instances" of the concept are physically more similar to each other than stimuli randomly assigned to each category. This factor accounts for the transfer to new exemplars as well as the unifying effect of the rule.

Our approach is to treat concept formation simply as a problem in pattern recognition. However, many difficult problems remain to be solved. For example, our model cannot now deal with the findings of D'Amato and Van Sant that transfer was based upon patches of red in photographs that contained much other detail. What principles lead to the selection of these aspects of the photographs? This and a number of related problems must be solved before it is possible to say whether a theory in which transfer is carried by physical similarity can account fully for the process of "concept formation" in animals.

Applications of the Heinemann-Chase Model to Diffuse Stimuli

Our pattern recognition model is part of a general model of memory and decision making. The model was originally designed to account for generalization involving simple stimuli, such as the intensity of a light or sound. I will therefore start with a description of such an
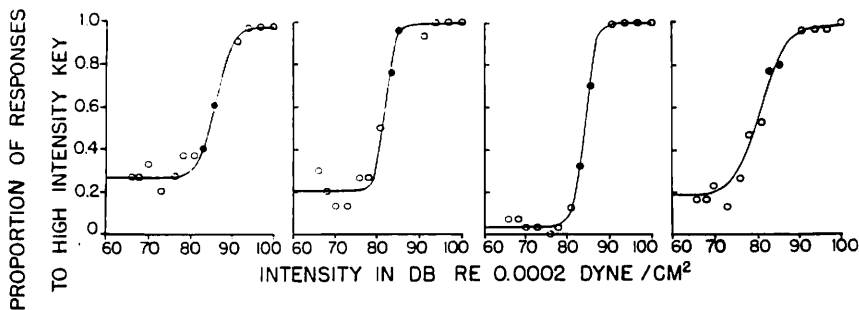
Fig. 2. Distribution of choices obtained during generalization test following training to discriminate between levels of white noise differing by 2.3 dB. The black dots show the proportion of responses to the high intensity key in the presence of the training stimuli. Each panel shows the results for one pigeon.

experiment.

In this experiment (Heinemann. *et al.,* 1969) pigeons were trained to identify, by pecks on the appropriate response key, two sounds that differed only in intensity. They were rewarded with food for making one response, R1, when presented with the softer sound, and an alternative response, R2, when presented with the louder sound. For one group the two sounds differed by only 2.3 dB. After prolonged training the pigeons were presented with 11 sound intensities in addition to the two used in training. During this generalization test they tended to chose the key that had been correct (during training) for the softer sound when presented with any sound of intensity less than or equal to that of the less intense training stimulus. Similarly, they chose the key correct for the more intense training stimulus for all stimuli of intensity more than or equal to that of the training stimulus. One could describe this choice behavior as reflecting application of the concepts "soft" and "loud" to the new stimuli. These data are shown in Figure 2. Note that the proportion of R2 responses rises gradually with sound intensity, following a curve that is similar to the psychometric function one would obtain from humans tested for intensity discrimination with the method of single stimuli. The birds acted as though they had learned the "rule": make R1 for soft sounds and R2 for loud sounds.

I will present a brief account of how our model handles these data. For a fuller account see Heinemann (1983a; 1983b), Chase and Hei-

nemann (1990) or Chase and Heinemann (1991).

Our theoretical account of these data assumes two stages in the acquisition of a discrimination. There is a *presolution period* (PSP), the length of which is dependent upon the difficulty of the discrimination. During the PSP the subject is assumed to perform an analysis designed to discover significant statistical associations between information arriving over any of its sensory channels and outcomes of behavior, e.g. reward or non-reward. The formal model is based on the Sequential Probability Ratio Test (Wald, 1947). With respect to the present experiment, it is assumed that the observer discovers that the sensory channels over which information predictive of reward or non-reward arrives are the auditory channels that transmitted the band of frequencies contained in the white noise used in the experiment. I shall refer to these channels as "relevant channels." Only information arriving over relevant channels will be processed during the second stage of learning.

The second stage of learning involves a memory that has a large but not unlimited storage capacity. This memory will be referred to simply as the *long-term memory* (LTM). On each trial of an experiment the subject is assumed to deposit in the LTM a record that contains:

(a) *A representation of the stimulus information transmitted over the relevant channel(s).* In the example under consideration, the particular loudness experienced when the choice was made is represented as a point on a loudness

continuum. The physical differences in the a-coustic stimuli determine the separations be-tween the corresponding points along the loud-ness continuum.

(b) *A representation of the response made.* For the analysis described here this is represented by a label, e.g. R1 or R2.

(c) *A representation of the reward received.* For the analysis described here the records are la-beled as positive if food was delivered; negative if it was not.

Each record is said to occupy a *storage location* in LTM. The location to which each record is sent is selected randomly and any record occu-pying a storage location will be destroyed ("ove-rwritten") when a new record is entered at that location.

It is assumed that during each trial the sub-ject retrieves from the LTM a small random sample of positive records (records showing that a reward was received). Each sample con-tains from 3 to 18 records. The only informa-tion that is used in the response selection pro-cess is the information contained in this sample of records in working memory.

The choice of response is based on a compari-son of the current sensation, the *current input,* to the information in working memory. It is as-sumed that memory of the previous sensory experiences is imprecise—the remembered sen-sation is distorted by Gaussian noise—small dis-tortions are frequent, large distortions rare. Af-ter a record has been retrieved, and is being held in working memory, the sensation represented on that record fluctuates rapidly over time, mo-mentary values falling into a Gaussian distribu-tion whose mean represents the value retrieved from the LTM. Each of the four distributions shown in Fig. 3 represent a record of a remem-bered sensation. In this illustration three of the records show that R2 was rewarded and one shows that R1 was rewarded.

The response the subject selects is the one that the retrieved information indicates is most likely to earn a reward. To find this response the subject gets the sum of the probability den-sities for each response at the current input, and selects the response for which the sum of the densities is the greatest. The process amounts to summing the heights of the R2 curves above the
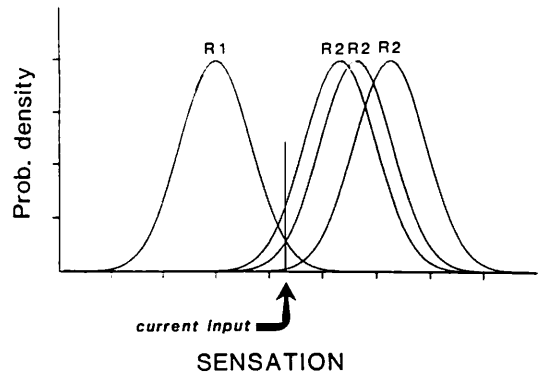

Fig. 3.　A sample of four records retrieved from LTM. The choice of response is based on the probability densities at the point labeled current input.

point representing the current input, doing the same for the R1 curves, and making the re-sponse for which the sum is highest. Errors will occur—not only is the memory of the sensation imprecise, the decision is based on a small num-ber of remembered events. Nonetheless, given these limitations, this decision rule optimizes the probability that the decision made is cor-rect.

If the probability density at the current input is below some very small threshold value, a new sample is drawn. This rarely happens when the stimuli vary along a single dimension. How-ever, in the case of more complex stimuli, such as the pattern of stimulation produced when a picture is viewed, all records in the sample may be very different from the current one. After having retrieved a sample, the subject retrieves another sample if, and only if, the sample in hand provides no information about the conse-quences of past behavior in the presence of sensations reasonably similar to the one cur-rently experienced. Our estimates suggest that, if approximately ten attempts to retrieve useful information have failed, the subject simply chooses the response associated with the largest expected value of reward. Using this simple notion we have been able to account for the marked improvement in identification of stimu-li as they increase in complexity (Chase & Hei-nemann, 1989).

Our model provides the following explanation

for the generalization data shown in Fig. 2: After training, the LTM is filled with records showing that R1 was rewarded in the presence of a sensation that was induced by training-stimulus S1, and R2 was rewarded in the presence of a sensation induced by training-stimulus S2. On the average, sensations induced by S2 will be remembered as more intense than sensations induced by S1, but the distributions of remembered intensities overlap. Because random samples retrieved from LTM reflect the distribution of records in LTM, it will be true of the sample also that the records showing remembered R1s will be associated with remembered sensations that are less intense than those associated with remembered R2s.

As illustrated in Fig. 3, the decision rule specifies that the response made is the one for which the probability density at the current input is greatest. Probability densities for R1 and R2 tend towards identity at the category boundary, the point bisecting the distance between the two training stimuli. As the distance from the category boundary increases, the expected difference in the summed probability densities for the two alternatives increases, resulting in curves which are monotonic with lower and upper and lower asymptotes of 0 and 1.0. Deviations from these asymptotes are attributed to "guessing." Guessing occurs when repeated sampling has failed to produce useful stimulus information (i.e. a sample yielding an above-threshold probability density at the current input.)

Applications of the Heinemann-Chase Model to Patterned Stimuli

In situations in which spatial information is irrelevant, as is true for diffuse stimuli such as sound intensities, we treat sensory information as though it arrived over a single "sensory channel," the channel that the presolution period has shown to be relevant. In extending our model to visual pattern recognition we assume that at any moment in time the continuum of sensations induced by a pattern of stimulation may be represented by a sample of evenly spaced "sensory channels." Patterns, such as the dot matrix figures produced by illuminating pixels on a computer screen, are represented simply by
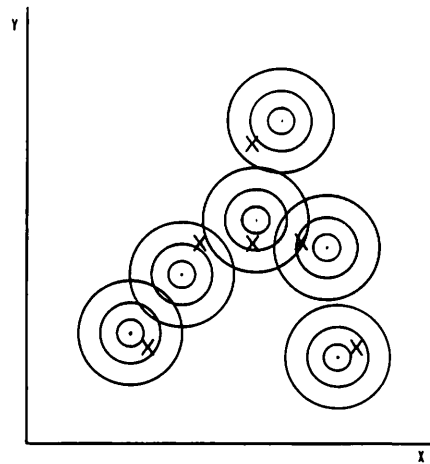

Fig. 4. Remembered group of points representing the letter A as shown on a single record retrieved from LTM. The concentric circles represent contours of constant probability density on the bivariate distributions for the spatial coordinates of each point. The x's represent points on the current input.

the location of the pixels (the X and Y coordinates) defining the pattern. It is assumed that while a record of a visual pattern resides in LTM, the values of the X and Y coordinates of each point vary randomly and independently over time, the distributions of momentary values being Gaussian. A record of this type is illustrated in Figure 4. The concentric circles represent the bivariate density functions representing the remembered letter A. The x's represent the current input points produced by the dot matrix letter A.

According to our model, a subject trained to recognize a number of different dot-matrix letters when shown a particular letter for identification, retrieves a few records from LTM and compares the current input pattern to each of the patterns represented on the retrieved records in order to find the best match. We assume that response selection is based on a decision quantity D, computed as follows: At each point on the current input calculate the mean probability density contributed by each point on the memory record. (A mean density that falls below the threshold is assigned a value of

zero.) This will yield as many means as there are current-input points. The decision quantity, D, is equal to the product of these means. If two or more records in the sample represent the same response, the value of D associated with that response is the sum of the individual D values. The decision rule is: make the response associated with the largest value of D. If no record is associated with a value of D that is greater than zero, draw a new sample. Finally, if successive samples fail to yield a non-zero value of D for any record, then choose the response associated with the greatest probability of reward in the past. This rather complicated-sounding procedure is simply an extension of the decision rule described earlier for the treatment of unidimensional stimuli. It is a type of "fuzzy template" matching scheme in which the current input (a pattern) is compared to remembered patterns as illustrated in Figures 3 and 4). This model is described in more detail in Heinemann and Chase (1990).

Through computer simulation we have been able to demonstrate the power of our model as a predictor of pigeons' performance on recognition of dot matrix figures. We used the model to simulate choice behavior in a pattern recognition experiment involving the 26 letters of the alphabet. The confusion matrix generated showed that certain errors, e.g. confusions between E and F will occur frequently while others, e.g. A and T rarely occur. The simulated matrix compared quite well with one published by D. Blough (1985) for confusions made by pigeons and with confusion matrices for human observers obtained by Podgorny and Garner (1979) using either reaction times or similarity ratings. Simulation of some of Cerella's (1990) work on rotational invariance is reported in Heinemann's paper for this symposium. Contextual effects in pattern recognition have been examined in papers by Donis and Heinemann (1993) and Donis. *et al.*, (1994). The simulations which come closest to providing evidence for transfer of a "concept" followed the procedures used by Van Hamme. *et al.*, (1992) in an experiment that was designed to test the recognition-by-components (RGB) theory of Biederman (1987). In their experiment pigeons identified four outline drawings—a pigeon, a turtle, a rolling pin and a lamp

—with partially deleted contours. When the original contours were replaced by the deleted segments there were decrements in performance, decrements that closely matched those generated by our computer simulation but were not compatible with RGB theory. Our model also generates confusion matrices showing the probabilities with which various incorrect responses will be made when an error occurs. Comparison of such matrices to pigeon generated ones could lead to a stronger test of our theory but, unfortunately, pigeon-generated confusion matrices were not available for the experiment we analyzed.

The colored photographs used in most concept formation experiments are far different from the small dot matrix patterns that we have used in our simulations. Does the animal subject remember every detail of the entire photograph? How large an area of the retinal image is processed? Data such as those obtained by Yamashita (1991) suggest that only a small part of the retinal image is processed —a few degrees of visual angle—if the discriminative stimuli are imaged on a fixed retinal region. However, the region is much larger if the discrimination requires attention to a larger region. Yamashita's data show that pigeons can process stimuli over a range of about 30 degree of visual angle horizontally to the left and right of a fixation point. Some of Yamashita's data are shown in Heinemann's paper for this symposium.

Conclusions

1. Use of photographs depicting complex scenes as discriminative stimuli, introduced by Herrnstein and Loveland in 1964, has raised new and important questions regarding the processing of information by non-verbal animals. For example, is all the information available in the pictures processed and, if not, what determines which aspects are? Are the pictures seen as representing the three-dimensional, constantly changing, world or are they meaningless to the subjects? To what extent does performance on these categorization tasks reflect past experience of the subject or the species?

2. The term "concept" is too vague and ill defined to serve as an explanatory construct. What is needed now is careful experimentation

designed to uncover the mechanisms responsible for the behavior attributed to the formation of a concept.

3. The Heinemann-Chase model of pattern recognition gives a satisfactory account of many of the known facts of stimulus categorization by animals but, thus far, is unable to deal with misclassifications that seem to reflect the perceptual organization of the extended visual stimulus.

REFERENCES

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

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.

Cerella, J. (1990) Shape constancy in the pigeon: The perspective transformations decomposed. In M. L. Commons, R. J. Herrnstein, S. Kosslyn & D. B. Mumford (Eds), *Behavioral Approaches to Pattern Recognition and Concept Formation* (Vol. 8, pp. 145–163), 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. 4, pp. 37–57), Cambridge, MA: Ballinger

Chase, S., & Heinemann, E. G. (1989) Effects of stimulus complexity on identification and categorization. *The International Journal of Comparative Psychology, 3*, 165–181.

Chase, S., & Heinemann, E. G. (1991) Memory limitations in animal and human signal detection. In M. L. Commons, J. A. Nevin & M. C. Davison. (Eds), *Signal Detection: Mechanisms, Models and Applications.* (pp. 121–138). Hillsdale, NJ: Erlbaum

D'Amato, M. R., & Van Sant, P. (1988) The person concept in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 43–55.

Donis, F. J., & Heinemann, E. G. (1993) The object-line inferiority effect in pigeons. *Perception and Psychophysics, 53*, 117–122.

Donis, F. J., Heinemann, E. G., & Chase, S. (1994) Context effects in visual pattern recognition by pigeons. *Perception and Psychophysics, 55*, 676–688.

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. 4, pp.

209–229). Cambridge, MA: Ballinger.

Heinemann, E. G. (1983a) 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. 4 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. 4, pp. 21–35). Cambridge, MA: Ballinger

Heinemann, E. G., Avin, E., Sullivan, M. A., & Chase, S. (1969) An analysis of stimulus generalization with a psychophysical method. *Journal of Experimental Psychology, 80*, 215–224.

Heinemann, E. G., & Chase, S. (1990) A quantitative model for pattern recognition. In M. L. Commons, R. J. Herrnstein, S. Kosslyn & D. B. Mumford (Eds.), *Quantitative analysis of behavior: Computational and Clinical Approaches to Pattern Recognition and Concept Formation.* (Vol 9, pp. 109–126). Hillsdale, NJ: Erlbaum.

Heinemann, E. G., & Chase, S. (1995) A quantitative model for simultaneous brightness induction. *Vision Research, 35*, 2007–2020.

Herrnstein, R. J., & Loveland, D. H.(1964) Complex visual concept in the pigeon. *Science, 146*, 549–551

Herrnstein, R. J. (1979). Acquisition, generalization, and discrimination reversal of a natural concept. *Journal of Experimental Psychology: Animal Behavior Processes, 5*, 118–129.

Podgorny, P., & Garner, W. R. (1979) Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. *Perception and Psychophysics, 26*, 37–52.

Van Hamme, L. J. Wasserman, E. A., & Biederman, I. (1992) Discrimination of contour-deleted images by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 18*, 387–399.

Vaughan, W., & Greene, S. L. (1983) In. M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative Analyses of behavior: Discrimination processes*, (Vol. 4, pp. 37–57), Cambridge, MA: Ballinger

Vaughan, W., & Greene, S. L. (1984) Pigeon visual memory capacity. *Journal of Experimental Psychology, 10*, 256–271.

Wald, A. (1947). *Sequential Analysis.* New York: Dover Publications

Yamashita, H. (1991) Frontal visual field Position and visual discrimination learning in the pigeon (*Columba livia*). Doctoral disseration. City University of New york.