Title
Categorization and the Parsing of Objects

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 16(0)

Authors
Pevtzow, Rachel
Goldstone, Robert L.

Publication Date
1994

Peer reviewed

Categorization and the Parsing of Objects

Rachel Pevtzow **Robert L. Goldstone**
rpevtzow@ucs.indiana.edu rgoldsto@ucs.indiana.edu
Department of Psychology
Indiana University
Bloomington, IN 47405

Abstract

Several models of categorization suggest that fixed inputs (features) are combined together to create categorization rules. It is also possible that categorization influences what features are perceived and used. This experiment explored the possibility that categorization training influences how an object is decomposed into parts. In the first part of this experiment, subjects learned to categorize objects based on particular sets of line segments. Following categorization training, subjects were tested in a whole-part decomposition task, making speeded judgments of "does whole X contain probe Y." All diagnostic and nondiagnostic category parts were used as parts within the whole objects, and as probes. Categorization training in the first part of the experiment affected performance on the second task. In particular, subjects were faster to respond when the whole object contained a part that was diagnostic for categorization than when it contained a nondiagnostic part. When the probe was a diagnostic category part subjects were faster to respond that it was present than absent, and when the probe was a nondiagnostic part, subjects were faster to respond that it was absent than that it was present. These results are discussed in terms of perceptual sensitivity, response bias, and the modulating influence of experience.

Introduction

In many traditional models of categorization, a fixed set of primitive features is used to determine categories. However, it is not obvious how these features are defined. Perhaps we perceive the lowest level of properties into which an object can be decomposed (such as line segments), or maybe we perceive some higher level (such as angles formed by line segments). One possibility is that the categories we use determine what the primitive inputs are. If this is the case, then learning, particularly learning new categories, may influence what features are perceived.

Evidence exists which suggests that, to some extent, we learn how to perceive; that is, our perceptions are influenced by experience. Gibson and colleagues (Gibson, 1969; Gibson & Walk, 1956; Gibson & Gibson, 1955) discussed ways in which perceptual learning can occur. One of these methods, pre-differentiation, involves increased ability to differentiate objects as a result of experience with them. In one study (Gibson & Walk, 1956), a group of rats were exposed to circles and triangles, but were not required to respond to these shapes. These rats were later able to learn to discriminate between these shapes much easier than a control group of rats which did not have this experience. Pre-differentiation of the shapes increased performance in a task which involved distinguishing between them. Another form of perceptual learning involves detection of distinctive features (Gibson, 1969). In this type of learning, features which are diagnostic of a particular object are learned. These features may not have been immediately distinguished, upon first experience with the object. Through experience, however, the observer learns which features can be used to determine the object's identity. These features may be created from more primitive features. In this case, learning can be said to be diagnosticity driven. An example of diagnosticity driven learning comes from an experiment performed by Waller (1970). A group of rats were trained to discriminate between floors of different colors and textures. These rats, trained to differentiate between a rough, black floor and a smooth, white floor, were later able to learn a new discrimination task that required them to distinguish between rough and smooth floors faster than a control group that received no prior training. In this study, the experimental group demonstrated diagnosticity driven learning. As a result of the training, they were able to learn to discriminate between different textures in a later transfer task.

Diagnosticity driven learning was further investigated by Goldstone (in press). He examined the effect of categorization on perceptual sensitivity to single feature dimensions, such as brightness and size. In these experiments, the issue was whether or not people can become more sensitive to dimensions that are diagnostic for a categorization task. One manner in which categorization was found to influence perceptual sensitivity was through acquired distinctiveness. In one experiment, the stimuli consisted of squares of varying size. Size differences between any two adjacently sized squares were scaled such that the perceived difference between any two adjacently sized squares were equal. A group of subjects underwent categorization training using size as the diagnostic feature, with Objects 1 and 2 belonging to one category and Objects 3 and 4 belonging to the other category. Subjects receiving this categorization training were later able to discriminate between Objects 2 and 3 better than those who did not receive this categorization training. In addition, subjects receiving categorization training became more perceptually sensitive to size differences between categories than within categories, such that they were better at detecting size differences between Objects 2 and 3 than between Objects 1 and 2.

This study examined perceptual learning along single stimulus dimensions. However, it does not necessarily imply that the same type of learning can occur for more complex features, involving multiple stimulus dimensions. Is it possible for something varying along multiple dimensions to be treated as a single feature? Czerwinski, Lightfoot, & Shiffrin (1992) demonstrated that complex stimuli, conjunctions of values along multiple dimensions, can become unitized through extensive practice. They found another type of perceptual learning: perceptual unitization of multiple dimensions, to form a single feature.

One difference between experts and novices seems to be perceptual unitization. For example, people who are chess experts divide the board into features differently than do novices (Chase & Simon, 1973). They found that whereas a novice may perceive only a random collection of pieces, an expert may perceive an ordered arrangement. This is analogous to Czerwinski et al.'s work on perceptual unitization. The individual pieces and their locations can be likened to basic stimulus dimensions. The experts create and perceive the complex features based on their unitization of the dimensions. This unitization may allow them to extract more information from the board than novices can.

There is some evidence that people can become sensitized to complex features. Experts perceive structures in x-rays (Norman, Brooks, Coblentz, & Babcook, 1992), beers (Peron & Allen, 1988), and infant chickens (Biederman & Shiffrar, 1987) that are missed by novices. As people learn to distinguish between and categorize objects, they learn which features are relevant for these tasks. One demand of these tasks is to learn to perceive diagnostic features which may not be immediately apparent to absolute novices.

Parsing and Part Decomposition

There is frequently more than one way to parse an object into its component parts. Palmer (1977, 1978) argued that some parts are inherently better than others, and that this parsing will naturally occur in such a way as to create the best parts. He developed a set of converging measures designed to gauge the "goodness" of a part. One of these involved part-whole verification: Palmer assumed that as goodness of a part increases, less time would be necessary to identify whether or not the part was present in the whole. For example, in Figure 1, subjects saw the whole object on the left, and one of the four parts on the right. It usually took subjects less time to confirm that one of the good parts was contained in the whole than that one of the poor parts was. In another method, subjects performed a "mental synthesis" task. In this task, subjects were given two non-overlapping parts, similar to the ones in Figure 1. They were asked to create a whole object, by imagining the two parts overlapping. After they completed this task, they were shown a whole object and asked whether it was the same as the object they mentally constructed. In this task, it was assumed that better parts would be synthesized more quickly and accurately than poorer parts. A third method used subjects' ratings of the goodness of parts, and the last method examined how subjects spontaneously parsed objects. All four methods converged, showing that the good parts in Figure 1 are better than the poor parts. Palmer created a model to compute a rating of goodness of parts. The criteria used in this model include connectedness and continuity of two lines, as well as the location, length, and orientation of individual lines. Using his formula, he was able to model the obtained data with considerable accuracy.

Palmer's model assumes that the goodness of a part is completely constrained by stimulus factors. It computes part goodness using only characteristics of the whole and part. It does not say anything about the possible influence of subject factors on goodness ratings of parts. Palmer did not say that subject factors were unimportant, but they were not a part of his model. However, it may be possible for subject factors to influence the measured goodness of a part, such that the empirically observed best parts would differ from Palmer's calculated goodness values. As discussed above, it may be possible that people can learn features, and become more perceptually sensitive to them. If this is true, the rating of goodness may change as a function of the experimental training without modifying the part or whole in any way. For example, using a part-whole decomposition task, subjects may be able to respond more quickly to features if they have become more sensitive to them. Palmer used response time in a part-whole decomposition task as one measure of a part's goodness.

If categorization influences perceptual sensitivity and object parsing, then Palmer's model cannot be complete, although it would still be valuable. In this experiment, we tested his model by examining the influence of categorization on performance in a whole-part decomposition task. We used this task because it is largely a perceptual test. Subjects underwent categorization training, learning two 3-line segment parts which were diagnostic for the categories. Subjects were divided randomly into two categorization groups, with each group learning different diagnostic parts. Both groups then participated in a whole-part decomposition task, determining whether a given probe was present within the whole. Response latencies of subjects in different categorization conditions were compared, to determine how diagnosticity influenced performance during the decomposition task.

There are a couple of ways that categorization training might affect performance in the whole-part decomposition task. It is possible that perceptual learning might cause a general increase in sensitivity to diagnostic parts. If this were the case, then we would expect faster response times

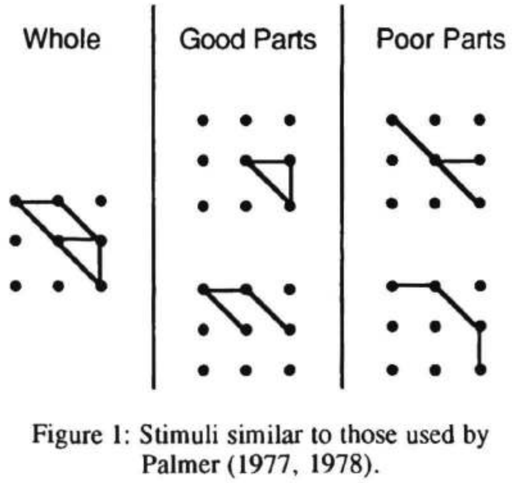

Figure 1: Stimuli similar to those used by Palmer (1977, 1978).

when the whole object contains a diagnostic part, or when the probe is a diagnostic part. Another possibility is that perceptual learning might induce a response bias. Subjects might be inclined to respond "present" when a diagnostic part is present in the display, and "absent" when a nondiagnostic part is present (or vice versa). Finally, some combination of these effects might be found.

Method

Subjects

The subjects were 45 undergraduate students at Indiana University. They received class credit for their participation.

Apparatus

This experiment was conducted using Macintosh IIsi computers. Subjects were seated 42cm from the display.

Materials

Stimuli were made up of line segments connecting dots on a 3 by 3 grid (Figure 2). The distance between any two horizontally or vertically adjacent dots was approximately 2cm. Diagonal line segments were approximately 2.5cm long. The dots and line segments were black, presented on a white background.

Procedure

There were two tasks in the experiment: categorization and whole-part decomposition. In the categorization phase of the experiment, subjects were shown distortions of Objects 1, 2, 3, 4, 5, 6, 7, and 8 as shown in Figure 2. Distortions were created by adding a black line segment at a random location. Subjects were asked to place an object into one of three categories. In one categorization condition, Objects 1 and 2 belonged to one category, with A as the defining feature, and Objects 3 and 4 belonged to the other category, with B as the defining feature. In this condition, A and B were diagnostic because they were used to discriminate between two categories, while A' and B' were nondiagnostic. In the other categorization condition, Objects 1 and 3 belonged to the same category (A'), and Objects 2 and 4 belonged to the other category (B'). In this condition, A' and B' were diagnostic for categorization, and A and B were nondiagnostic. Subjects were randomly assigned to the two categorization conditions. Objects 5, 6, 7, and 8 were always members of the third category. Subjects responded using the 'z' and '/' keys and the spacebar to indicate whether the object presented was a member of category 1, 2, or 3, respectively. Following the response, a check was displayed if the subject was correct, and an 'X' appeared if the subject was incorrect. The stimulus was presented in approximately the center of the display. It remained on the screen for the entire trial, and was cleared 500 ms after the feedback was presented.

In the second phase of the experiment, trials consisted of displays with 'wholes' and 'probes'. The wholes consisted of one of the four category-defining features (A, B, A', or

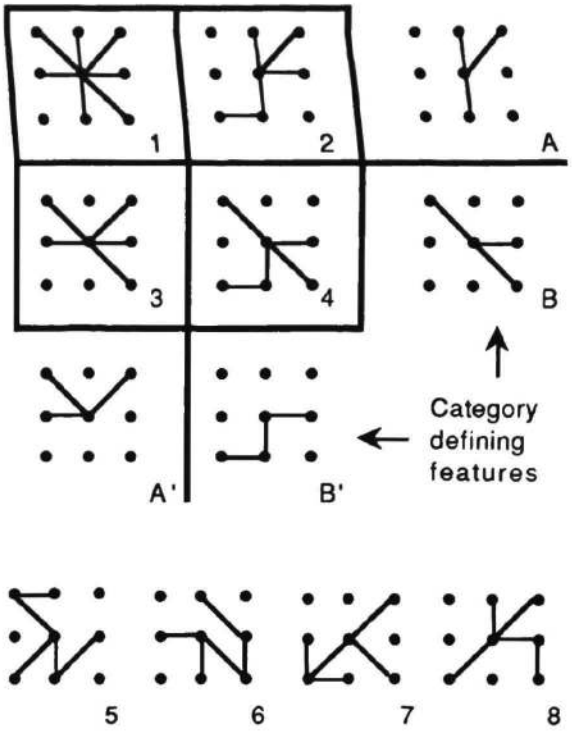

Figure 2: stimuli for categorization phase of experiment

B') from the categorization task, plus three connected line segments (complements), which were connected to the category part. The complements were constrained by the rule that they could have no lines overlapping with any of the category parts. They served as a control for the category parts. There were a total of 36 possible complements, producing 144 whole objects. The probes were either category-defining features, or complements.

There were four types of trials in whole-part decomposition task: present category probe, absent category probe, present complement, and absent complement (Figure 3). For each of the trials shown in Figure 3, the object on the left is the whole, and the object on the right is the probe. In the first type of trial, the probe is a category part which is contained within the whole object. In the second, the probe is the complement to the category-defining part. In the absent category part trials, the probe is a category part, but is not contained within the whole object. For the last type of trial, absent complement, the probe is a randomly chosen complement from another object. On these trials, the probe and whole would have between zero and three line segments in the same (relative) locations.

Wholes were presented alone for 1000ms, and then a probe was added to the display. The wholes were displayed in the same location as the object during the categorization training, and the probes were placed 4.5cm to the right and .5cm down from the right, top edge of the whole, displaced about 5cm to the right of the whole. The subjects' task was to decide, as quickly and accurately as possible, whether or not the whole contained the part. Subjects responded 'present' by pressing 'z', and '/' was used for 'absent' responses. Both stimuli remained visible until the

Table 1: Conditions in whole-part decomposition task

Type of Probe	Category part in whole object	Probe	Presence of Probe within whole object	number of trials
Category	Nondiagnostic	Nondiagnostic	absent	12
Category	Nondiagnostic	Nondiagnostic	present	24
Category	Nondiagnostic	Diagnostic	absent	12
Category	Diagnostic	Nondiagnostic	absent	12
Category	Diagnostic	Diagnostic	absent	12
Category	Diagnostic	Diagnostic	present	24
Complement	Nondiagnostic	*not applicable*	absent	24
Complement	Nondiagnostic	*not applicable*	present	24
Complement	Diagnostic	*not applicable*	absent	24
Complement	Diagnostic	*not applicable*	present	24

subject responded. After the response was made, subjects were told whether or not they were correct.

In the whole-part decomposition task, there were four stimulus properties of interest: whether the probe is a category part or a complement;whether the whole contains a diagnostic part; whether the probe is a diagnostic category part; and whether or not the probe is contained within the whole. The ten conditions formed by these properties are presented in Table 1. These conditions form a complete set of logically possible combinations of the relevant properties.

The tasks were blocked: subjects performed a block of categorization training, followed by a block of whole-part decomposition. The blocks alternated, and were repeated three times each. Before each block, subjects were shown the instructions for the task. In order to continue to the decomposition task, subjects had to meet a minimum criterion of 90% accuracy over 24 trials of categorization. Subjects were told about this requirement, and received feedback regarding their accuracy rate after every 24 trials. In addition, in the first block of categorization training, subjects performed 48 practice trials. Thus, each subject participated in at least 120 categorization trials. Category 1, 2, and 3 trials occurred equally often. The experiment took approximately an hour to complete.

Results and Discussion

Only trials in which the subject responded correctly were included in the analyses. Trials with response times faster than 100ms and slower than 10 seconds were excluded. Figure 4a shows the mean response times to decide whether or not the part was present in the whole, as a function of whether or not the whole contained a diagnostic category part. When the probe is a category part, response times are faster when the whole object contains a diagnostic category part than a nondiagnostic part. This advantage is seen for both present and absent category parts, although it is greater when the probe is part of the whole. A different pattern is seen when the probe is a complement. In this case, present probes elicit faster responses when the category part in the whole is nondiagnostic than when it is diagnostic, while absent probes are associated with faster response latencies when the whole contains a diagnostic part than when it contains a nondiagnostic part.

Response times to respond to category parts were faster ($F(1,44)=3.93$, $p=.05$) for wholes containing a diagnostic category part than for those containing an nondiagnostic part. This diagnosticity advantage was significant only for present category parts ($t=2.02$, $p<.05$, $df=44$), although a similar trend was also seen on absent category part trials ($t=1.11$, $p>.05$, $df=44$).

For complements, there was a main effect of presence ($F(1,44)=5.73$, $p<.05$), with faster response latencies to present than absent complements. Unlike category parts, however, there were no main effects of diagnosticity ($F(1,44)=.84$, $p>.05$), nor was there an interaction of diagnosticity and presence ($F(1,44)=1.13$, $p>.05$).

While Figure 4a showed response times as a function of the diagnosticity of the category parts within the whole, Figure 4b shows the mean response times of category probes as a function of whether or not they were diagnostic. That is, in Figure 4a, it was the parts within the whole that were diagnostic and nondiagnostic, while in Figure 4b it is the probes themselves that were diagnostic and nondiagnostic. As can be seen, diagnosticity again influenced response latencies. However, the influence of diagnosticity varied, depending on whether or not the probe was present. When the probe was absent, probes that were nondiagnostic during categorization training had an advantage over ones that were diagnostic. On the other hand, times were faster for diagnostic present probes than nondiagnostic present probes.

There was a significant interaction of diagnosticity and presence of the probe ($F(1,44)=6.61$, $p<.05$). Diagnostic category probes show an advantage for response latencies ($t=2.02$, $p<.05$, $df=44$) over nondiagnostic category probes. On the other hand, when subjects have to decide that a diagnostic probe is not present, there is a slight tendency to be slower to respond than when the probe is nondiagnostc ($t=-1.2$, $p>.05$, $df=44$).

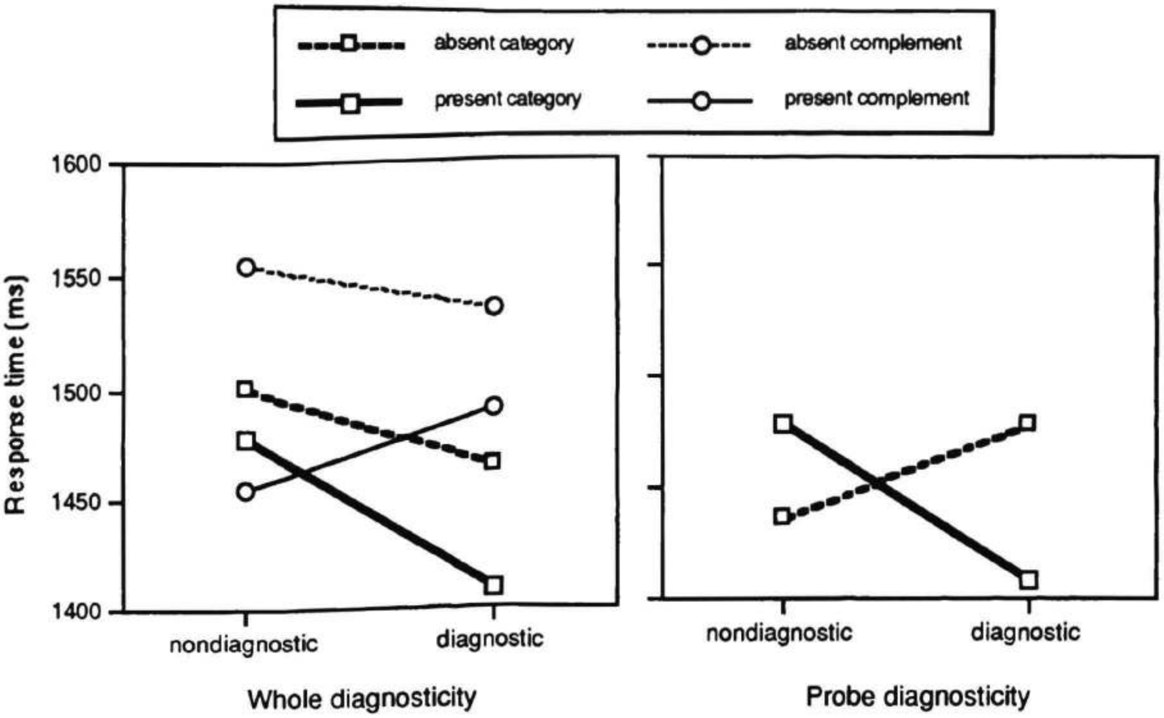

Figure 4: Response times graphed as a function of diagnosticity of category part within whole (A), and as a function of diagnosticity of probe (B).

There were main effects of part type ($F(1,44) = 10.31$, $p<.01$) and presence of part ($F(1,44)=5.93$, $p<.05$). Subjects were, in general, quicker to respond to category parts than complements. They were also faster to respond "present" than "absent."

Conclusions

Subjects underwent categorization training, learning two parts that were diagnostic for the categories. One group of subjects used parts A and B (from Figure 2) as the diagnostic parts, while another group used A' and B'. When these two groups were compared, it was found that subjects were able to respond more quickly when the whole object contained a diagnostic part than when the object contained an nondiagnostic part. In one method used by Palmer (1977, 1978) to measure a part's "goodness", response times in a part-whole decomposition task were compared for different parts. His conclusion was that better parts were associated with lower response times. Using this same method we can conclude that because diagnostic parts are associated with faster response times than nondiagnostic parts, diagnostic parts acquire some extra amount of goodness. Thus, we can conclude that Palmer's model is incomplete. It is necessary to incorporate subject factors into this model, as well as stimulus factors.

When the whole object contained a diagnostic part, subjects were faster to respond than to wholes containing nondiagnostic parts. We used two categorization groups. The only thing that differed for these groups was whether the parts were diagnostic; both groups saw the same stimuli. According to Palmer's model, the two categorization groups should have the same response times for any given whole/probe judgment. An increase in perceptual sensitivity to diagnostic parts could explain the difference in "goodness." When subjects see a diagnostic part within an object, the part may "pop out" as a feature. As a result, when subjects parse the whole object it is more natural for the diagnostic part to become one of the resulting parts. According to Palmer's model, an object is naturally parsed into its best parts. Therefore, it seems that diagnostic parts increase their "goodness" when compared to nondiagnostic parts.

Signal detection theory is a way of assessing how people respond to information. There are two response dimensions, sensitivity and response bias. Sensitivity refers to the ability to know when to respond, and when not to respond. Response bias, on the other hand, is a tendency to respond in a particular way. Both of these dimensions could be subject to learning. We use the terms "sensitivity" and "bias" in a manner analogous to their use in signal detection theory. In this experiment, sensitivity is the ability to correctly respond "present" and "absent." Bias is the tendency to respond "present" or "absent."

Response bias can partially explain the results found for the presented category probes. As with complements, responses to present diagnostic probes are faster than present nondiagnostic probes, and absent nondiagnostic probes elicit faster responses than do absent diagnostic ones. It seems that when a diagnostic probe is presented, there is a bias to say "present", regardless of whether it's actually present within the whole. Likewise, subjects showed a bias to respond "absent" to nondiagnostic probes. If the correct response is consistent with the bias, subjects are fast to respond. However, if the correct response goes against their bias, it takes longer to respond.

Both a change in perceptual sensitivity and a response bias were found. Subjects increased their sensitivity to diagnostic parts. Regardless of the probe, subjects were able to respond more quickly when the whole object contained a diagnostic part. Subjects viewed the whole object for 1000ms before the probe was presented. This time would give them the chance to mentally parse the object into its best parts. Parts which were diagnostic during the categorization training may stand out as the best parts, allowing subjects to respond faster than when the whole contains an nondiagnostic part. The response bias, on the other hand, was determined by the probe. When the probe was a nondiagnostic category part, subjects were faster to respond "present" than "absent." Conversely, when the probe was a nondiagnostic category part, subjects were faster to respond "absent" than "present." So, sensitivity changes with the diagnosticity of the category part within the whole, while diagnosticity of the probe affects response bias.

The parsing of an object into its best component parts is influenced by both stimulus and subject factors. The stimulus factors defining "goodness" of parts, as described by Palmer, include Gestalt properties such as connectedness. However, this model does not completely determine how an object is parsed. The categorization training influenced perceptual sensitivity and response bias to parts, and thus changed their "goodness." The concepts we create influence sensitivity not only to our preexisting features, but also to features that we have not yet created.

Acknowledgements

Many useful comments and suggestions were provided by Dedre Gentner, John Kruschke, Douglas Medin, Robert Nosofsky, Philippe Schyns, Richard Shiffrin, Edward Smith, Linda Smith, Keith Smith, and Jean-Pierre Thibaut. We thank Kristen Best, Sandy Conklin, Matt Golden, Sarah Hoffman, and Tanya Stamm for their assistance in conducting the experiment. This research was funded by Biomedical Research and Support Grant PHS 307 RR7031N from the National Institute of Mental Health. Correspondance concerning this paper should be addressed to rpevtzow@ucs.indiana.edu or Rachel Pevtzow, Department of Psychology, Indiana University, Bloomington, Indiana 47405.

References

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 640-645.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

Czerwinski, M., Lightfoot, N., & Shiffrin, R.M. (1992). Automatization and training in visual search. The American Journal of Psychology, 105, 271-315.

Gibson, E.J. (1969). Principles of perceptual learning and development. Appleton-Centure-Crofts, New York.

Gibson, E.J., & Walk, R.D.(1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. Journal of Comparative and Physiological Psychology, 49, 239-242.

Gibson, J.J. & Gibson, E.J. (1955). Perceptual learning: Differentiation of enrichment? Psychological Review, 62, 32-41.

Goldstone, R.L. (in press). Influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General.

Norman, G. R., Brooks, L. R., Coblentz, C. L., & Babcook, C. J. (1992). The correlation of feature identification and category judgments in diagnostic radiology. Special Issue: Memory and cognition applied. Memory & Cognition, 20, 344-355.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. Cognitive Psychology, 9, 441-474.

Palmer, S. E. (1978). Structural aspects of visual similarity. Memory & Cognition, 6, 91-97.

Peron, R. M., & Allen, G. L. (1988). Attempts to train novices for beer flavor discrimination: A matter of taste. The Journal Of General Psychology, 115, 403-418.

Waller, T.G. (1970). Effect of irrelevant cues on discrimination acquisition and transfer in rats. Journal of Comparative and Physiological Psychology, 73, 477-480.