On the Unitization of Novel, Complex Visual Stimuli

Nancy Lightfoot and Richard M. Shiffrin
Indiana University
Bloomington, IN 47405
lightfoo@ucs.indiana.edu
shiffrin@ucs.indiana.edu

Abstract

We investigated the degree to which novel conjunctions of features come to be represented as perceptual wholes. Subjects were trained in a visual search task using novel, conjunctively-defined stimuli composed of discrete features. The stimulus sets were designed so that successful search required identification of a conjunction of at least two features. With extended training, the slope of the search functions dropped by large amounts. Various transfer tasks were used to rule out the possibility that the organization of sequential search strategies involving simple features could account for this result. The perceptual discriminability or confusibility of the stimuli exerted an important influence on the rate of unitization. The nature of the perceptual unit appears to depend on the subset of features which are diagnostic for carrying out a particular discrimination task. The results provide important constraints for models of visual perception and recognition.

What are the mechanisms by which representations of individual features are bound together and processed as perceptual wholes? One solution to this binding problem involves pre-specifying a representation for all possible conjunctions of features, in addition to the features themselves (Hummel & Biederman, 1990). This approach has been criticized, however, as being relatively inefficient and implausible when applied to higher level conjunctions (Hummel & Biederman, 1990; Hinton, McClelland & Rumelhart, 1986).

This research was supported by grants AFOSR90-0215 to the Institute for the Study of Human Capabilities, and by grants NIMH 12717 and AFOSR 870089 to the second author.

We conducted a number of experiments to examine the nature of processing for conjunctively defined stimuli, and the way in which such stimuli might become unitized during training. Our specific interest was in examining the transition from the processing of unfamiliar stimuli at the level of individual features, to the processing of those same stimuli, after familiarization, as conjunctive perceptual wholes. Our results suggest that perceptual unitization occurs after prolonged exposure to particular stimuli. These findings argue against the notion that conjunctive representations are pre-specified. In conditions under which conjunctive representations have developed, however, they may form the basic functional units on which attentional processes operate.

The processing of visual information is often investigated using visual search tasks, in which the subject attempts to locate a target among multiple distractors, typically responding 'target present' or 'target absent'. The slope of the function relating response time to display size provides a convenient way of assessing capacity demands associated with increasing stimulus and display complexity. This slope is often assumed to reflect the time taken for a single comparison between a subject's representation of the target and a stimulus appearing in the display. Slopes on positive (target present) trials are typically half as steep as slopes on negative (target absent) trials. This pattern of results has been interpreted as evidence for a self-terminating search process, since subjects will on average identify the target half-way through their search of the display on positive trials. The increase in reaction time as a function of display size is assumed to reflect the operation of a limited-capacity search mechanism.

Many well-known theories of visual search assume that stimuli are processed at the level of

primitive visual features such as line orientation, curvature, or color (eg. Triesman & Gelade, 1980; Fisher, 1986). Treisman and her colleagues (eg. Treisman & Gelade, 1980; Treisman & Gormican, 1988), for example, propose that subjects first parse the visual field into individual feature maps, in which the presence of visual features is coded without information as to their location, or as to the objects to which they belong. According to Treisman, it is only through limited-capacity attentional processing that features are conjoined into coherent objects.

Triesman argues that the distinction between pre-attentive processing, in which the simple presence of features is coded, and attentive processing, in which these features are conjoined into coherent wholes, accounts for differences in efficiency in visual search tasks. Targets which are distinguishable from distractors on the basis of a single primitive feature will be detected automatically and without capacity limitations, whereas targets which require identification of a conjunction of features to be uniquely identified will require attentional processing.

An important limitation of Treisman's approach is its difficulty in accounting for the well-documented effects of training in visual search tasks. Two types of training paradigms have been studied extensively in visual search: Consistent mapping (CM) and varied mapping (VM) training (eg. Shiffrin & Schneider, 1977; Schneider & Shiffrin, 1977). In CM training, stimuli are designated either "targets" or "distractors" and never change roles across trials. In VM training, on the other hand, stimuli change roles randomly from trial to trial, appearing as targets on some trials, and distractors on others. In some search tasks, there is a large advantage for CM training, which is well accounted for by automatic attention attraction to targets (Shiffrin & Schneider, 1977; Schneider & Shiffrin, 1977). In other visual search tasks, however, typically where search is more difficult, slopes appear to decrease consistently across days of training in both CM and VM training paradigms, although there is generally at least some degree of CM advantage.

Fisher (1986) has proposed a feature-based model which accounts for a wide pattern of training effects found in visual search (see Shiffrin & Schneider, 1977; Schneider & Shiffrin, 1977 for an alternative explanation). According to Fisher's model, stimuli are decomposed and processed at the level of individual features, as Treisman has suggested. In visual search tasks, subjects sequentially compare the individual features of a target with those of the distractors. Display items not having the first feature are rejected, remaining items are tested on the next feature, and so on, until search terminates.

In Fisher's (1986) model, as in Treisman's model, targets distinguishable on the basis of a primitive feature will be easily identified. In the case of conjunctively defined targets, however, search efficiency depends on the particular order in which the features are compared. With increasing experience in searching for a given target among all of its possible combinations of distractor elements, subjects learn to organize a maximally efficient feature comparison strategy. The development of this search strategy gives rise to the observed effects of training.

There are several critical implications of feature-based theories. If all stimuli are assumed to be processed at the level of individual features, and all stimuli require limited-capacity resources to be conjoined into coherent wholes, then the basic units of information and the nature of processing should be roughly similar for stimuli with high and low levels of familiarity. This same prediction holds in the case where conjunctions of features form the basic unit of processing, but these conjunctive representations are pre-specified within the visual processing system. With regard to training effects, if stimulus sets are designed in which all feature-based search strategies require approximately the same number of comparisons to uniquely identify the target, there should be greatly attenuated effects for training.

Experiment 1

We tested these assumptions by training subjects on sets of novel stimuli in which featural overlap was carefully controlled. Figure 1 shows the two stimulus sets assigned to each subject. For each subject, one set was assigned to CM training and one to VM training. The stimuli were composed of an external frame and three internal line segments. The stimulus sets were designed so that within each set, all feature comparison sequences would lead to approximately comparable performance when averaged over displays composed of different distractor elements.

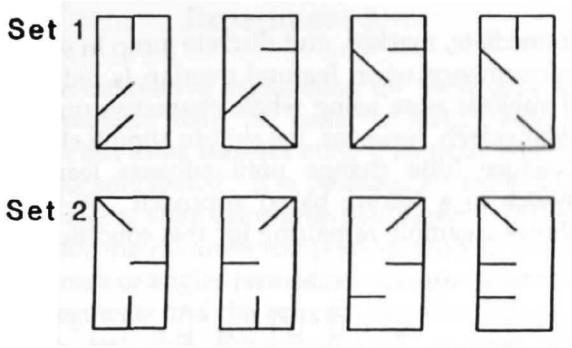

Figure 1. Novel stimulus sets for Experiment 1.

We assumed that the three internal line segments within each character formed the functional features in the search task. Based on this assumption, each target shared a feature with exactly one other stimulus in the set. Each target could be uniquely identified only by examining a conjunction of two features, but any two features would, averaged across displays of different distractor compositions, work equally well in identifying the target. There was no featural overlap between the two stimulus sets.

Under these conditions, training effects would be unlikely to be due to learning optimal feature comparison strategies. Furthermore, to the extent to which all stimuli are decomposed and processed at the level of individual features, performance on these characters should be comparable to that for familiar stimuli. As a result, extended practice using these stimuli should produce greatly attenuated training effects under the assumptions of feature-based models of visual search.

Subjects were run in CM and VM conditions for a period of 50 days. In the CM condition, one stimulus was designated as the target and the other stimuli were designated as distractors over the entire course of training. In the VM conditions, all stimuli within the set appeared equally often as targets, and served as distractor elements on remaining trials. In order to equate training on particular combinations of targets and distractors in the two training conditions, there were four times as many trials in VM as in CM training. Half of the trials in each condition were positive and half were negative. Display sizes varied from one to eight. For display sizes larger than three, displays were filled in a pseudo-random manner, with as few repetitions of each distractor stimulus as possible.

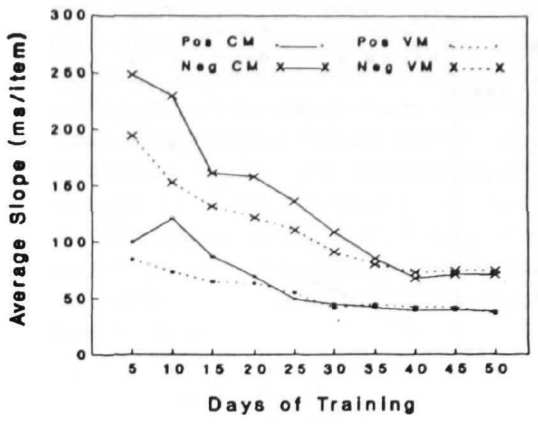

Figure 2. Training effects for Experiment 1.

Figure 2 shows the average slopes across days of training for positive and negative trials, using CM and VM training paradigms. These data are remarkable in several respects. First of all, the training effects are unusually large, with comparison times dropping from a high of 240 to an asymptote of 80 msec per item in the negative CM condition. The effects of training are also unusually protracted. Search performance does not asymptote for 35 to 40 sessions using these novel stimuli. A third unusual aspect of the data is that performance is initially better for VM than for CM training trials. Although CM and VM training can lead to comparable performance in visual search, an advantage for VM training is unprecedented.

The huge effects for training found with these novel stimuli suggest that there is an important role for familiarity in the processing of visual stimuli. This hypothesis is also supported by the fact that performance on CM trials was initially worse than on VM trials. Because there are four times as many trials in VM as in CM, a plausible explanation for this finding is that subjects are simply gaining more familiarity with the VM stimuli early in training. This argument is strengthened by the finding that the training paradigm itself has little influence on asymptotic search performance: reaction time functions at asymptote, graphed in Figure 3, are virtually identical for the two training conditions.

The nature of these familiarity effects is somewhat open to question. One explanation is that through repeated exposure, subjects are learning to unitize these novel stimuli and treat them as perceptual wholes. If subjects learn to perform comparisons at the level of the entire

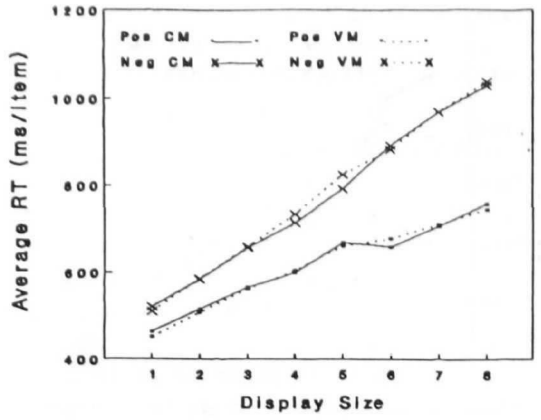

Figure 3. Display size effects at asymptote.

stimulus, instead of at the level of individual features, this greatly reduces the number of comparisons required to identify a target. This reduction in the actual number of comparisons offers a plausible explanation for the large training effects found in Experiment 1.

An alternative possibility relies on the high similarity among the characters within each search set. Both the large degree of featural overlap and the addition of the redundant external frames made the stimuli within a set highly similar, and guaranteed that search would be extremely difficult. It is possible that under these conditions, subjects were initially confused and grossly inefficient in developing appropriate feature search strategies for these displays. Subjects may have initially done a great deal of re-checking, for example, or may have attended to irrelevant or non-diagnostic features, such as the line segments in the external frame, before organizing a more coherent search strategy.

Experiment 2

To further investigate the nature of these familiarity effects, we trained the same subjects who had participated in our first experiment in an additional search task, in which the featural overlap of the targets and distractors was reduced. In particular, we made the items from the former VM set consistent targets, and the distractors from the former CM set consistent distractors. Targets and distractors now shared no features, so a feature based search should always terminate with the first comparison. Thus, if subjects were using a feature based search in Experiment 1, we should see an immediate, marked, and discrete jump in search performance when featural overlap is reduced. If subjects were using whole characters units in their search, however, the shift in stimuli should produce little change until subjects learn to switch to a feature based approach. Figure 4 shows a sample re-pairing for this condition.

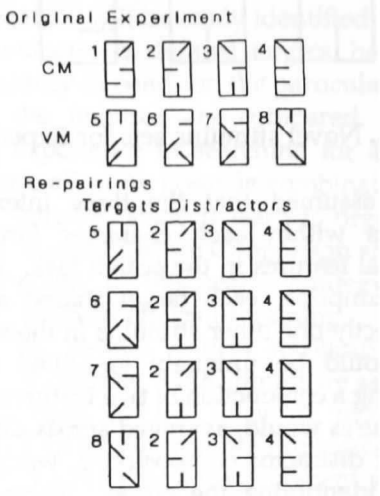

Figure 4. Sample stimulus sets, Experiment 2.

Subjects trained using these new stimulus sets for 15 sessions. Figure 5 shows five days of baseline performance using the original stimulus sets, followed by the data from the stimulus re-pairing. The results are not very compatible with the notion that subjects were employing a feature-based search. There was no sudden jump in performance after re-pairing. Instead slopes showed a moderate and continuous decline over the full course of training in the new condition, possibly reflecting a gradual switch to a more efficient feature-based search.

Figure 5. Results of stimulus re-pairing.

Experiment 3

An alternative explanation for these data is that subjects may be using a feature-based search, but using features other than those which we had anticipated. It is possible, for example, that subjects were using emergent features as the basis for their feature comparison process, such as corners or angles formed between the internal line segments and the external stimulus frames.

To test this hypothesis, we decided to continue training on the original search sets with the external stimulus frames removed. Most models of visual search would predict an improvement in search performance after removing the external frames, since search is much more efficient when the featural overlap between targets and distractors is reduced. If, on the other hand, subjects are relying on emergent features as the basis of their search strategy, removing the external stimulus frames should lead to an elimination of the emergent features and a disruption in search performance.

Figure 6 shows the slopes for five days of baseline training followed by the removal of the external stimulus frames. To establish a baseline, the subjects from the first two experiments were re-trained using the original stimulus sets (conjunctively-defined, framed stimuli). After re-training, the external frames were removed, and search with the unframed stimuli began.

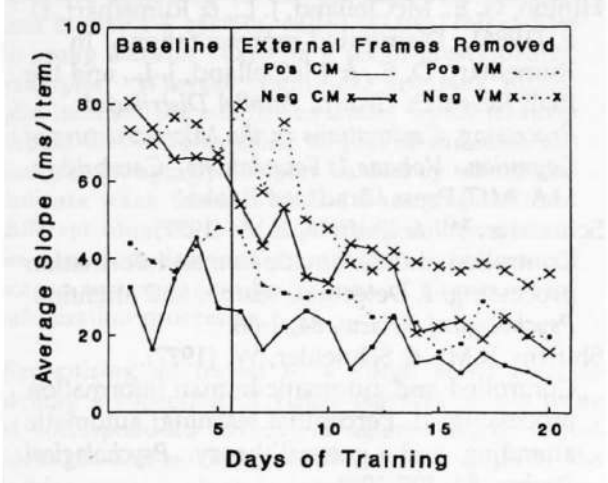

Figure 6. Results of removing external frames.

There was no evidence for disruption in search performance following the removal of the external frames. In fact, search gradually improved, as would be expected based on the decreased similarity between targets and distractors after the removal of the redundant line segments. These data offer no support for the idea that subjects are using emergent features as the basis for a strategic feature search. Given this demonstration that the internal stimulus features appear to form the basis for the search process, the results of Experiment 2 suggest that this process operates at the level of the character or feature conjunction, rather than at the level of individual features.

Experiment 4

This last study raises a puzzle, however. If search is based on unitized character representations, why does removal of a large part of the character not disrupt performance? One possibility is that the functional representation of these characters does not include the boundaries. Perhaps a good deal of the training involves learning to ignore the redundant and confusing external frames. If the learned unit consists of the arrangement of internal line segments, then the large degree of transfer seen in Experiment 3 would be expected. If this reasoning is correct, then initial training on unframed characters should be fast and easy, but subsequent transfer to framed stimuli should be very poor. We therefore trained new subjects in just this way. Experiment 1 was repeated using unframed stimuli. When subjects reached asymptote, the frames were added and training continued. The results are shown in Figure 7.

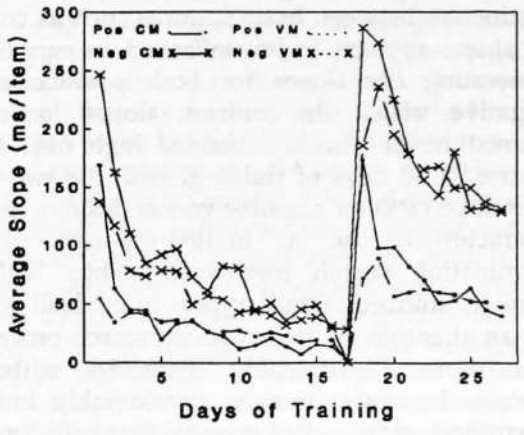

Figure 7. Slopes for subjects trained initially on unframed stimuli, before and after the addition of external stimulus frames.

The results are remarkable: Slopes began at the level indicative of feature search, but dropped quickly to a low asymptotic level. When the frames were added, no transfer was seen -- performance reverted to a level at least as poor as that seen at the start of training in both this experiment and in Experiment 1. This was followed by an additional gradual reduction in slopes, presumably reflecting the subjects' learning to remove the external frames from their perceptual representations.

This complete asymmetry of transfer, depending on the order in which the external frames are added or deleted, further reveals the nature of perceptual unitization. The learned units are apparently not the items themselves, but rather the parts of the items that are useful for making required discriminations.

Taken together, our data provide evidence for a process of perceptual unitization with increasing exposure to novel stimuli. Although it is logically possible to generate a feature-based explanation for the huge training effects seen in Experiment 1, this explanation is incompatible with the results of Experiments 2 and 3. However the exact nature of this perceptual unitization is somewhat ambiguous. Subjects may be learning to unitize the character as a whole (though without the external frames), or they may be learning to unitize only a simple diagnostic conjunction of two features.

It seems clear that these perceptually unitized representations do not act very much like the primitive visual features which Treisman has proposed as the functional units in perceptual processing. Search based on the distinction between basic features such as color or shape appears to be unlimited in capacity, generating flat slopes for both positive and negative trials. In contrast, slopes for our framed novel stimuli remained high over the course of 50 days of training, with the two-to-one slope ratio for negative versus positive trials characteristic of a limited-capacity, self-terminating search mechanism. Thus highly similar, unitized stimuli appear to be dealt with by an attentive, limited-capacity search process. The more discriminable characters without frames, however, showed considerably lower asymptotic slopes. It is possible that sufficiently discriminable conjunctions may come to exhibit characteristics similar to those for simple features.

We are clearly able to encode and recognize novel combinations of features, though with some difficulty, probably due to the necessity of processing such stimuli one feature at a time. With increasing familiarity with conjunctions of features, however, we apparently develop an alternative form of conjunctive or unitized representation of whole perceptual objects. Evidence for this type of perceptual unitization calls into question both the assumption that familiar stimuli are processed simply at the level of basic features, and the assumption that conjunctions of features are somehow pre-specified in the visual system. The fundamental differences in the processing of stimuli based on primitive visual features, novel combinations of features, and familiar visual wholes place strong constraints on models of visual processing and perceptual binding.

References

Fisher, D. L. (1986). Hierarchical models of visual search: Serial and parallel processing. Paper presented at the annual meetings of the Society for Mathematical Psychology, Cambridge, MA.

Hummel, J. E., & Biederman, I. (1990). Dynamic Binding in a Neural Network for Shape Recognition. Ph.D. Diss., Dept. of Psychology, University of Minnesota, Minneapolis, MN.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In Rumelhart, D. E., & McClelland, J. L., and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press/Bradford Books.

Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Triesman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15-48.