# Unitization During Category Learning

Robert L. Goldstone
Indiana University

Correspondence Address: Robert Goldstone
Department of Psychology
Indiana University
Bloomington, IN. 47408

Other Correspondences: rgoldsto@indiana.edu
(812) 855-4853

Running head: CATEGORIZATION AND UNITIZATION

Abstract

Five experiments explored the question of whether new perceptual units can be developed if they are diagnostic for a category learning task, and if so, what are the constraints on this unitization process? During category learning, participants were required to attend either a single component or a conjunction of five components to correctly categorize an object. Evidence consistent with unitization was found in that the conjunctive task became much easier with practice, and this improvement was not found for the single component task, or for conjunctive tasks in which the components could not be unitized. Influences of component organization (Experiment 1), component contiguity (Experiment 2), component proximity (Experiment 3), and number of components (Experiment 4) on practice effects were found. Using a Fourier transformation method for deconvolving response times (Experiment 5), prolonged practice yielded responses that were faster than predicted by an analytic model that integrates evidence from independently perceived components.

## Unitization During Category Learning

Interest in the mutual interactions between our perceptual and conceptual systems has resurfaced repeatedly. In anthropology, Sapir and Whorf (Whorf, 1941/1956) posited influences of language on perception, and in psychology, the “New Look” movement (Bruner & Postman, 1949) argued that high-level cognitive processes altered people’s ability to identify objects, make discriminations between objects, and make accurate judgments of an object’s attributes. Both of these historical movements have been resurrected, as new evidence for linguistic relativity (Gumperz & Levinson, 1991) and a “New New Look movement” (Niedenthal & Kitayama, 1994) have emerged.

The research reported here explores one possible influence of conception on perception: the functional unitization of perceptual components due to categorization experience. It is generally acknowledged that our categorization judgments are driven by our perceptions, and this influence is incorporated in almost all successful models of categorization (Nosofsky, 1986). However, it is possible that our categorizations also drive our perceptions. In particular, if a certain stimulus component is diagnostic for a required categorization, people can increase their categorization accuracy by becoming selectively tuned to the component. If the component is already part of the person’s perceptual “vocabulary,” then categorization demands can be met simply by increasing attention to the component (Kruschke, 1992; Nosofsky, 1986). However, if the component does not already exist in one’s vocabulary, then becoming tuned to the component first entails adding it to the vocabulary (Goldstone & Schyns, 1994; Schyns, Goldstone, & Thibaut, in press; Schyns & Murphy, 1994).

“Perceptual vocabulary” refers to the elementary building blocks used to create object representations. Evidence for membership in a perceptual vocabulary is typically based on functional behavior. For example, Treisman and Gelade (1980) argue for a vocabulary of features that includes color, orientation, and closure elements. These features are empirically identified by response times to targets that are not influenced by the number of distracters if the targets are distinguished by a feature, fast segregation of textures if the textures have unique features (Julesz, 1981), illusory conjunctions involving features from different objects if attention is not focused on the objects, and all-or-none disappearance of features when images are retinally stabilized (Hebb, 1949). Evidence for vocabulary elements can also be obtained by neurophysiological recordings, such as Hubel and Wiesel’s (1968) original work indicating primary visual cortex neurons selectively tuned to particular line orientations (for more recent work, see Perrett & Oram, 1993).

The above research is based on the implicit assumption that a fixed vocabulary of features can account for our perceptual experience. In fact, some support for this assumption comes from results showing little influence of practice in augmenting feature vocabularies. For example, Treisman & Gelade (1980) found that prolonged experience in a conjunctive feature search task (searching for a blue “O” among blue “T”s and red “O”s) did not result in a significantly reduced dependency of response times on the number of distracter elements, suggesting that the conjunction of color and form features was not added as a separate vocabulary element over time. However, other research has shown an influence of life-long (Wang, Cavanagh, & Green, 1994) and laboratory-induced (Laberge & Samuels, 1974; Shiffrin & Lightfoot, 1997; Treisman, Vieira, & Hayes, 1992) experience on feature search. The research to be reported here further tests the adequacy of this “fixed vocabulary” hypothesis, not within a feature search paradigm, but in a categorization task. The particular type of vocabulary additions that I will explore are unitizations - situations where a single “chunk” is formed from perceptual components if the chunk is diagnostic.

## Perceptual Consequences of Category Learning

The notion that category learning can have perceptual consequences has recently been the source of substantial empirical inquiry. Work on categorical perception has shown that our ability to make perceptual discriminations is influenced by the categories we possess (see Harnad, 1987). Specifically, discriminations involving pairs of stimuli that straddle a category boundary are more easily made than are discriminations involving stimuli that fall within the same category, equating for physical dissimilarity between the pairs. Further research has indicated that the visual categories that influence perceptual discriminability need not be innate, but rather may be learned within a laboratory context (Goldstone, 1994; Lane, 1965). Laboratory-acquired visual categories can also influence perceptually-based similarity ratings (Harnad, Hanson, & Lubin, 1994). Objects that are placed in different categories are given lower similarity judgments than the same items are when they are placed in the same Livingston, Andrews, & Harnad, 1998; Kurtz, 1996). Lin and Murphy (1997) have shown that the conceptual knowledge associated with a category, manipulated by assigning different functional descriptions to the same items, can influence participants’ speed at verifying perceptual properties. Likewise, Jonides and Gleitman (1972) found that the interpretation of an ambiguous symbol influenced how quickly visual searches were conducted when it was a target. In short, learned categories can influence the speed and sensitivity with which perceptual properties are processed.

In addition, category learning can affect how an object is segmented into parts. Schyns and Murphy (1993, 1994) formulated a Functionality Principle whereby if a fragment of a stimulus is diagnostic for categorization (distinguishes members from non-members), then the fragment is instantiated as a unit. Consistent with this principle, they found that objects were more likely to be segmented into parts that were useful for categorizing them (see also Schyns & Rodet, 1997). Subjective segmentations were obtained by asking participants to draw outlines around the parts of objects. A similar influence of categorization on the segmentation of objects was shown by Pevtzow and Goldstone (1994). Stick figures composed out of six lines were categorized in one of two ways. Different arbitrary combinations of three contiguous lines were diagnostic for the different categorizations. After categorization training, participants made part/whole judgments, responding as to whether a particular set of three lines (a part) was present in a whole stick figure. Participants were significantly faster to determine that a part was present in a whole when the part was previously diagnostic during categorization. Whereas previous research has focused on objective properties of an object (e.g. the proximities, similarities, and shapes of the line segments) that determine how people will decompose it into segments (Palmer, 1978), the above results indicate that the person’s experience also influences how they will segment an object into parts. Such an influence of experience on segmentation is predicted by Behrmann, Zemel, and Mozer’s (1998) neural network model that groups together stimulus components that often co-occur and preferentially uses these learned groups for interpreting subsequent stimuli (see Goldstone et al, in press, for an alternative computational model).

An indirect source of evidence indicating an influence of concept learning on perceptual units that are formed comes from the “inversion effect” (Diamond & Carey, 1986; Tanaka & Gauthier, 1997; Tanaka & Farah, 1993; Yin,1969). According to this effect, the recognition cost of rotating a stimulus 180 degrees in the picture plane is much greater for specialized, highly practiced stimuli than for less specialized stimuli. For example, recognition of faces is substantially less rapid and accurate when the faces are inverted. This large difference between upright and inverted recognition efficiency is not found for less highly practiced objects. Diamond and Carey (1986) found a large inversion cost for dog breed recognition, but only for dog experts. Similarly, Gauthier and Tarr (1997; see also Gauthier, Williams, Tarr, & Tanaka, 1998) found that large inversion costs for a particular nonsense object can be created in the laboratory by giving participants prolonged experience identifying the object. They conclude that prolonged experience with an object leads to a configural representation of it that combines all of its parts into a single, viewpoint specific, functional unit. The relation between large inversion costs for recognition and configural processing is not completely clear. In addition to showing strong inversion costs, faces also show evidence of configurality in that identifying a face part is facilitated by preserving the rest of the face's features (Tanaka & Farah, 1993). However, this advantage of identifying a part within a preserved whole rather than in isolation is not always increased with expertise (Tanaka & Gauthier, 1997) as might be expected if perceptual learning leads to greater unitization.

Unitization

The previously reviewed literature indicates that one result of category learning is to create perceptual units that combine stimulus components that are useful for the categorization. Such a process is one variety of the more phenomenon of unitization, by which single functional units are constructed that are triggered when a complex configuration arises (Goldstone, 1998). Cattell (1886) invoked the notion of perceptual unitization to account for the advantage that he found for tachistoscopically presented words relative to non-words. Gestalt psychologists proposed the perceptual principle that objects will tend to be perceived in terms of components that have acquired familiarity (Koffka, 1935). Weisstein and Harris (1974) found that briefly flashed line segments are more accurately identified when they are part of a set of lines forming a unitary object rather than an incoherent pattern. They interpreted this effect as showing that arrangements of lines can form configural patterns that are perceived before the individual lines are perceived.

Unitization has been explored in the field of attention. Using a task where participants decided whether or not two visual objects were identical, Laberge (1973) found that when stimuli were unexpected, participants were faster at responding to actual letters than to letter-like controls. Furthermore, this difference diminished as the unfamiliar letter-like stimuli became more familiar over practice. He argued that the shape components of often-presented stimuli become processed as a single functional unit with practice.

More recently, Czerwinski, Lightfoot, and Shiffrin (1992) have referred to a process of perceptual unitization in which conjunctions of stimulus features are "chunked" together so that they become perceived as a single unit. Shiffrin and Lightfoot (1997) argued that separated line segments can become unitized following prolonged practice with the materials. Their evidence comes from the slopes relating the number of distracter elements to response time in a feature search task. When participants learned a conjunctive search task in which three line segments were needed to distinguish the target from distracters, impressive and prolonged decreases in search slopes were observed over 20 hour-long sessions. These prolonged decreases were not observed for a simple search task requiring attention to only one component. In addition, when participants were switched from a conjunctive task to a simple feature search task, there was initially little improvement in search times, suggesting that participants were still processing the stimuli at the level of the unitized chunk that they formed during conjunctive training. The authors concluded that conjunctive training leads to the unitization of the set of diagnostic line segments, resulting in fewer required comparisons.

Research on letter and word perception provides a useful source of information on unitization because life-long experience with a language causes some quasi-arbitrary arrangements of letters to occur frequently while other non-word arrangements rarely if ever occur. The evidence on whether the letters that comprise a word are unitized together into a single functional unit is mixed. In evidence favoring unitization, researchers have found that word units can be processed automatically and interfere with other processes less than do nonwords (LaBerge & Samuels, 1974; O'hara, 1980; Smith & Haviland, 1972). Results have shown that the advantages attributable to words over nonwords cannot be explained by the greater informational redundancy of letters within words (Smith & Haviland, 1972). Disrupting the standard physical appearance of a target in a lexical decision task by mixing upper and lower case letters harms word judgments more than non-word judgments, but only for brief presentations (Allen, Wallace, and Weber, 1995), a result interpreted as suggesting an initial bias to process words holistically. Words are often times identified faster than letters within words (Johnson, 1975), and this advantage disappears when the word or pattern-level information can not be unitized (Johnson, Turner-Lyga, & Pettegrew, 1986). Increasing familiarity with words can lead to people having difficulty detecting letters within the words (Tao, Healy, and Bourne, 1997), consistent with the notion that a strong word-level unit interferes with a decomposition of this unit into features.

However, other researchers have argued that results suggesting that recognition processes occur at levels higher than individual letters. Paap, Newsome, and Newell (1984) argue that using mixed cased letters (as Allen et al do above) may slow performance not by disrupting word-level shape cues, but by slowing processing at the letter level. In a review of the literature on holistic, word-level processing and analytic, letter-level processing, Besner and Johnston (1989) conclude that analytic processing is a necessary component in word recognition models. Although models of analysis and unitization in word perception are directly relevant to the current research, word perception may differ from the type of novel object perception explored here. On the one hand, the units formed during a lifetime of word learning are potentially stronger than the units learned following a few hours of experience with a novel shape. On the other hand, the bias for analytic processing may be greater for words than for our novel "doodles" because the letters that comprise a word are clearly delineated and separable entities whereas the parts that comprise doodles are fused together and contiguous with each other. As such, the current studies have a similar objective to the above word recognition literature, to explore evidence for perceptual processes that respond to information at levels higher than the individual components, but words and novel objects may yield different results.

<u>A New Source of Evidence for Unitization</u>

The purpose of the current set of experiments is to test whether category learning can lead to stimulus unitization, and to explore the boundary conditions on unitization related to stimulus characteristics and amount of training. The experiments explore unitization, but from a somewhat different perspective from the work in attention. First, the experiments are primarily interested in the influence of category learning on unitization, under the hypothesis that a unit will tend to be created if the parts that comprise the unit frequently co-occur, and the unit is diagnostic for categorization. Second, a new technique for analyzing response times is developed that compares response time distributions from a conjunctive categorization task to the expected distribution based on analytic models that do not incorporate unitization. Evidence for unitization is obtained if the observed response time distribution for a conjunctive categorization task contains response times that are faster than predicted by analytic models that base conjunctive responses on several individual component judgments.

Whenever the claim for the construction of genuinely new units is made, two objections must be addressed. First, perhaps the unit existed in people's vocabulary before categorization training. The stimuli are designed to make this explanation unlikely. Each unit to be sensitized is constructed by connecting 5 randomly chosen curves. There are 10 curves that can be sampled without replacement, yielding 30240 (10 X 9 X 8 X 7 X 6) possible different units. As such, if it can be shown that any randomly selected unit can be sensitized, then an implausibly large number of vocabulary items would be required under the constraint that all vocabulary items are fixed and *a priori*. The second objection is that no units need be formed; instead, people analytically integrate evidence from the five separate curves to make their categorizations. However, this objection will be untenable if subjects, at the end of extended training, are faster at categorizing the units than would be expected by the analytic approach. Quantifying what “faster than expected” means is the main business at hand, and will not fully be addressed until Experiment 5.

## Experiment 1

Experiment 1 explores the unitization of visual components to attain greater efficiency in a categorization task. The categorization task is designed so that evidence from five components must be received before certain categorization responses are made. For this reason, the critical categorization is a conjunctive task. The stimuli and their categorizations are shown in Figure 1. Each letter refers to a particular segment of the stimulus. Each stimulus is composed of five segments, joined below by a broad U-shape in order to create a closed object. To correctly place the stimulus labeled “ABCDE” into Category 1, all five components, “A,” “B,” “C,” “D,” and “E,” must be processed. For example, if the right-most component is not attended, then ABCDE cannot be distinguished from ABCDZ which belongs in Category 2. Not only does no single component suffice for accurate categorization of ABCDE, but two-way, three-way, and four-way conjunctions of components (posited by researchers such as Gluck & Bower, 1988, and Hayes-Roth & Hayes-Roth, 1977) also do not suffice. For example, the three-way conjunction “C and D and E” is contained in the stimulus ABCDE, but this conjunction does not discriminate ABCDE from AWCDE or VBCDE. Only the complete five-way conjunction suffices to reliably categorize ABCDE.

If unitization occurs during categorization, then it is possible that the stimulus ABCDE comes to be treated functionally like a single component with training. If this occurs, then participants should be able to quickly respond that this stimulus belongs to Category 1. Instead of responding by integrating the results from separate components, the categorization response may eventually be made by consulting a single detector. As this detector unit becomes established, response times to categorize the ABCDE pattern should decrease. In the current experiment, a pronounced decrease in the time required to categorize the conjunctively defined stimulus ABCDE will be taken as evidence of unitization.

If an influence of practice on response time is in fact found, then model fits to a power function will be used to assess the nature of this influence. The power function takes the form

$RT = a + bN^c$, where N is the trial number, a is the asymptotic response time, b is the difference between the initial and asymptotic response times, and c is the learning rate (which controls the steepness of the learning curve). The power function has been widely used in a variety of tasks with generally good results (Logan, 1992; Rosenbloom & Newell, 1987). The constants a, b, and c will be fit to each condition of an experiment and statistical tests on these constants will assess differences between conditions. The unitization hypothesis as framed above is primarily concerned with the constant b. Unitization should produce markedly different response times before and after practice. To the extent that experimental factors can influence the total amount of improvement shown in a condition, differences in estimated values for b should be observed. The constant c is typically negative when fitting response times, indicating decreased rates of improvement with practice, and will be affected by unitization only if unitizeable stimuli affect the speed, rather than amount, of improvement. Similarly, variations in asymptotic response time indicate differences in the overall speed of response across trials. Differences between conditions on their asymptotic response times, but not b values, would suggest that conditions vary only in their overall ease rather than their susceptibility to improvement. As such, the unitization hypothesis most directly relates to b, although it is an empirical question to observe which factors do change across conditions.

In order for improvements in the conjunctive task to be taken as evidence for unitization, two important control conditions are necessary. First, it is important to show that tasks that do not require unitization do not show comparable speed ups. To this end, a control task is included that allows participants to categorize the item ABCDE by attending only a single component rather than a five-way conjunction. This "One" (component) condition should not result in as much speed up as the "All" (components) condition where all components must be attended. If it does, then the speed up can be attributed to a simple practice effect rather than unitization. Second, it is important to show that stimuli that cannot be unitized also do not show comparable speed ups. For this control task, a five-way conjunction of components must be attended, but the ordering of the components within the stimulus is randomized. As such, a single template cannot serve to categorize the ABCDE stimulus. Assuming that unitization requires that the components to be chunked appear in a consistent or template-like manner (Gauthier & Tarr, 1997; Schneider & Shiffrin, 1977), these randomly ordered stimuli should not afford unitization, and thus are not expected to yield significant speed ups.

In sum, Experiment 1 explores the possibility that five curved segments, when combined together to create a contiguous, coherent object, may become treated as a single unit with practice. It is highly unlikely that the unit existed in the participants' perceptual vocabulary prior to training because the unit is randomly selected from an extremely large set of other units (the set of all of possible ways of ordering five segments, each chosen from a set of ten segments). The potential evidence from this experiment for unitization would come from substantial improvement in a conjunctive task that involves unitizeable stimuli, but not in a conjunctive task involving hard-to-unitize stimuli, and not in a simple component detection task that does not require unitization.

## Method

Participants. Seventy-two undergraduate students from Indiana University served as participants in order to fulfill a course requirement. The students were evenly split into the four between-subject conditions.

Materials. Stimuli were formed by selecting five line segments without replacement from a set of 10 segments. The displayed objects consisted of five curved lines combined together, and joined by a bowl shape. Each horizontally arranged segment was 0.8 cm long and 0.5 cm high. The entire length of a stimulus was 4 cm, and the height was 1.7 cm. The viewing distance was approximately 40 cm, yielding a visual angle of 5.7 degrees for the horizontal length of each stimulus. The starting and ending points of each segment were located at the vertical midline, and consequently, all segments could be joined with each other to create a seamless stimulus. Sample stimuli are shown in Figure 1. Each of the ten segments can be associated with a different letter of the alphabet for descriptive purposes. The assignment of segments to letters was randomized for each participant. As such, the object that is abstractly represented by "ABCDE" was composed of different segments for different participants.

Design. In the "All" task, the objects that are abstractly described as "ABCDE" and "VWXYZ" belonged to Category 1, and the objects "ABCDZ," "ABCYE," "ABXDE," "AWCDE," and "VBCDE" belonged to Category 2. This stimulus structure is shown in Figure 1. Given these category memberships, all five curves need to be attended in order to reliably place the object ABCDE into Category 1. The five members of Category 2 were created by replacing one of ABCDE's segments with a new segment. The object ABCDE was presented four times more frequently than any of the other six objects, which were all presented equally often. Given this, the category validities (the probability of a segment, given a category) for every segment and category were equal, as were the cue validities (the probability of a category, given a segment). The item VWXYZ was included in Category 1 so that no segment, by itself, would provide probabilistic evidence in favor of Category 1 or 2.

In the "One" task, ABCDE and VWXYZ were in Category 1, but only one of the five objects in Figure 1's second category was presented. Thus Category 2 only contained one object, randomly selected from: ABCDZ, ABCYE, ABXDE, AWCDE, or VBCDE. The selected object was presented five times more frequently than it was in the All task, so that both categories were presented equally often. Given these categories, to categorize ABCDE as belonging to Category 1, it is only necessary to attend to a single segment. For example, if Category 2 contained the object ABCDZ, then noticing the presence of segment “E” provides sufficient evidence for a Category 1 judgment.

Procedure. Participants were given 320 categorization trials. On each trial, an object appeared on the screen, and participants pressed one of two keys to indicate the object’s category membership. The object appeared on the screen until a response was made. Participants received feedback indicating whether their guess was correct, and if incorrect, were shown the object’s correct categorization. Feedback was displayed for 400 msec, followed by an inter-trial blank interval of 400 msec. As such, there was no explicit warning signal given before the onset of each object, but participants could potentially use the 400 msec interval after the feedback to prepare for the next trial. Participants were told that both accuracy and speed were important and that they should make their responses as quickly as possible without sacrificing accuracy. Participants who made more than 5% errors on a particular block were asked to try to increase their accuracy. An equal number of Category 1 and 2 objects were displayed, randomly ordered. Each participant received a different random order.

Half of the participants were given objects generated according to the All task, and the other half were given objects generated according to the One task. Four groups of participants were obtained by combining this All versus One split with a second orthogonal split: Ordered versus Random. For the Random group, any ordering of the same components counted as the same object. As shown in Figure 2, any one of the 120 orderings of the five components A, B, C, D, and E acted as an example of ABCDE. In this condition, when an object was selected to be displayed, the spatial ordering of its five segments was randomized. In the Ordered condition, the spatial positions of the five segments was fixed. As such, the object ABCDE had the same exact appearance every time it was displayed in the ordered condition.

What object was displayed on a given trial was randomized, subject to the constraint that objects and categories were presented with their stipulated frequency over the course of the experiment. The spatial position of the object was randomized using a uniform distribution such that the object appeared within a 20 by 20 cm square positioned in the center of the screen. The 320 trials were broken down into 4 blocks of 80 trials. Rest breaks were provided between blocks. During the break, the average accuracy and response time for the preceding block of trials was displayed. In the initial instructions, participants were shown all of the objects and their correct category assignments, and viewed the objects until they felt they knew which objects belonged to which categories. At the end of the experiment, fifty-eight of the 72 participants (those participants that completed the experiment with sufficient time remaining to answer questions) were asked what strategies they used to categorize items as quickly as possible. In particular, they were asked which of the following two strategies best characterized their behavior: A) “I tried to look for each of the parts of the doodle that was important for categorizing it,” or B) “I tried to remember an overall image of what some of the doodles looked like, and used this to categorize them.” The experiment took about 55 minutes to complete.

## Results

Figure 3 presents the results of most interest. A 2 (Random vs Ordered display) X 2 (All vs One component necessary) X 4 (blocks) ANOVA was conducted with correct response time as the dependent measure. For this primary ANOVA, only response times for the object ABCDE were considered, because it is only this object that requires the full set of five segments to be attended in the All condition. A main effect of display was found, $F(1, 68) = 10.4$, mse = 15.2 (all mse values are reported in milliseconds), $p < .01$, such that Ordered displays were more quickly categorized than Random displays. A main effect of number of necessary components was found, $F(1, 68) = 18.3$, mse = 13.0, $p < .01$, with the One condition significantly faster than the All condition for the object ABCDE. A main effect of blocks was found, $F(3, 204) = 8.9$, mse = 18.2, $p < .01$, indicating that practice increased speed of categorization.

However, these main effects were modulated by a predicted three-way interaction between the ANOVA variables, $F(3, 204) = 9.1$, mse = 13.7, $p < .01$. As shown in Figure 3, the practice effect due to blocks was most pronounced for the All, Ordered condition. There was also a substantial practice effect for the All, Random condition, but it was not even half as large as it was for the All, Ordered condition. Practice effects for the two One conditions were quite small.

To assess the nature of the improvements across the conditions, power functions were fit to each participant's response times on each of the four blocks from each of the four conditions using non-linear regression within SPSS 6.0. That is, each participant provided separate estimates for the best-fitting values for the three constants of the power function. The average best fitting functions were:

$RT_{\text{All, Ordered}}\ 0.468 + 1.15\ N^{-0.942}$

$RT_{\text{All, Random}} = 0.491 + 0.71\ N^{-0.33}$

$RT_{\text{One, Ordered}} = 0.341 + 0.107\ N^{-4.62}$

$RT_{\text{One, Random}} = 0.450 + 0.316\ N^{-1.68}$

Predicted response times were measured in seconds and N refers to the block number (1-4). All four of these non-linear regressions account for at least 97% of the variance in response times as a function of practice. ANOVAs on the three constants indicated no influence of condition on the best-fitting value for constant a (asymptote), $F(3,204) = 0.2$, mse = 0.44, but influences of condition on both b (total improvement), $F(3, 204) = 6.1$, $p < .01$, mse = 0.29, and c (learning rate), $F(3, 204) = 3.2$, $p < .01$, mse = 1.06. Learning rate and total improvement are, if anything, negatively correlated, with steep learning rates belonging to the One conditions that show relatively small total amounts of improvement. One interpretation of the steep learning rates for c in the two One conditions is that general familiarization with the experiment leads to large improvements from Block 1 to Block 2, but thereafter, little improvement is observed because the simple stimuli offer little opportunity for increased processing efficiency. The most striking difference between the All, Ordered condition and the other conditions occurs on the b value.

In the All condition, response times for the other items were faster than for “ABCDE.” Response times for the five items belonging to Category 2 were combined because these items were logically equivalent. Overall, collapsing across the ordered and random groups of participants, the Category 2 items were responded to in 1195 msec, which was significantly different from the 1354 msec required to respond to ABCDE, paired $t(35) = 5.4$, $p < .01$. Similarly, although ABCDE occurred more frequently than VWXYZ, the average correct response time for categorizing the latter object, 1121 msec, was significantly faster, paired $t(35) = 4.9$, $p < .01$. Presumably, the relatively fast response times for this item is because any two-way conjunction of segments suffices to make the VWXYZ categorization, whereas ABCDE requires a full five-way conjunction in the All condition if a single unit is not formed and used. Thus, the slower response time for ABCDE than VWXYZ may reflect the imperfect unitization of ABCDE; even after practice some responses to this item may be made by combining evidence from individual components. Consistent with this hypothesis, the variances within individual participants' response times were much larger for ABCDE than VWXYZ, paired $t(35) = 7.1$, $p < .01$, suggesting that responses to ABCDE may reflect a mixture of responses based on units and explicitly combined information from individual components. Fast responses to ABCDE may reflect responses based on complex units, whereas slow responses reflect analytically combined evidence from simple components. Also relevant to the unitization hypothesis, the speed advantage of Category 2 items over ABCDE in the All condition decreases over blocks of practice, as shown by a significant Item X Block interaction, $F(3, 102) = 4.6$, mse = 1 3.2, $p < .01$.

The response time effects were mirrored by error analyses. Across the four blocks, the percentage of errors were 8.0%, 5.2%, 4.3%, and 4.0%, $F(3, 204) = 5.1$, mse = 1.8, $p < .01$, indicating improvement from the beginning of the experiment to the end. The error rates for the four conditions were: Random All = 8.8%, Random One = 4.7%, Ordered All = 5.2%, and

Ordered One = 2.8%, $F(3,204) = 8.4$, mse = 1.4, $p < .01$. These error rates exhibit the same ordering as the response times with the slowest task being the most error prone. In no case was there any indication of a speed-accuracy tradeoff where one condition yielded faster response times but decreased accuracy than another condition. There was no significant interaction between condition and blocks on error rates, $F(9, 612) = 0.7$, mse = 2.1, $p > .1$.

Participants in the four different conditions differed in their self-report of strategy. The percentages of participants claiming that they used the holistic strategy of memorizing entire objects, rather than looking for particular segments, were 25%, 7%, 75%, and 7% for the Random All, Random One, Ordered All, Ordered One conditions respectively, Chi-square (3) = 23.65, $p < .05$. This result is consistent with the response time evidence that participants in the Ordered All task develop a single unit to recognize the Category A item rather than decomposing it into parts and recognizing it through those parts. That is, participants in the Ordered All task, the task that was hypothesized to yield the strongest unitization, were the most likely to report basing their judgments on matching items to whole stored images of objects.

## Discussion

Experiment 1 provided suggestive evidence for unitization of a consistent, multi-segment stimulus in a categorization task. Unitization was suggested by the pronounced and gradual improvement in the speed of categorizing objects. This speed up was particularly striking when the object to be categorized required attention to be placed on all of its parts, and when the parts could be easily combined into a coherent image.

The first of these provisions is indicated by the far greater improvement in categorization for the All task than for the One task. In fact, after the first block of training, there was no significant improvement in the One task. This is predicted if the All task benefits from the construction of functional units that span across multiple segments, and if the construction of these units takes time. The gradual nature of unitization is supported by Czerwinski et al's (1992) observation that improvements in a conjunctive feature search task were found after even as much as 20 hours of practice.

The second provision, emphasizing the value of practice for unitizeable stimuli, is suggested by the comparison of the Ordered and Random conditions. The All task for both of these conditions requires attention to all five of the segments that compose a stimulus. However, when the five segments were randomly located within the stimulus, then practice effects were greatly attenuated. One explanation of this difference is that a single unit that represents an entire stimulus can only be formed when it is presented in a visually constant form. Expressed in a slightly different way, unitized representation may be photograph-like images in that they preserve the stimulus information in a relatively raw, unprocessed, spatially constrained form. This conjecture gains support from other work showing that units that are acquired over practice preserve many of the properties of actual images (Gauthier & Tarr, 1997; Shiffrin & Lightfoot, 1997). If units are constrained to be image-like representations, then randomization of the segments within an image would preclude unitization because it prevents a stimulus from being associated with a single image.

The power function modeling indicated that the unitizeable stimuli of the Ordered, All condition yield relatively large amounts of improvement in response times, rather than particularly fast overall response times or rates of improvement. Thus, the suggestion is that the hallmark of unitization should be the amount, rather than the speed, of improvement. This is consistent with the relatively gradual improvements that have been observed for conjunctively-defined targets (Shiffrin & Lightfoot, 1997).

The finding that practice effects were more pronounced for the Ordered All condition than for the Random All condition serves to eliminate alternative accounts of what cognitive process improved with practice. It might be thought that the stronger practice effects for the All, relative to One, task were due to segment comparison times that were facilitated by practice. By this reasoning, the All task required five times as many segment comparisons as did the One task; if practice made each comparison faster, then the All task would be expected to show five times the improvement of the One task. One problem with this account is that the practice effect for the Ordered All task was more than five times greater than it was for the Ordered One task. Also, the explanation cannot be this simple because the Random All task required at least as many segment comparisons as the Ordered All task did, but did not show a comparable practice effect. In fact, the Random All task required more comparisons if some segments that were consistently placed together in the Ordered condition came to act as a single segment. Furthermore, the Random task required cognitive processes not required by the Ordered task; namely, it required a process of segment localization, in which the relative location of a particular segment within an object was determined. In sum, the pronounced practice effects in the Ordered All condition were unlikely to have been due solely to sped up processes for identifying/comparing specific segments or integrating the results from separate segment comparisons; if they were, then the Random All condition would have shown comparable practice effects. Further, the pronounced practice effect was not due to developing faster processes for locating a segment in an object; if it were, then the Random All condition would have shown larger practice effects. This analysis further suggests that the practice effects observed for the Ordered All condition were due to a process that created a single, image-like unit. This conclusion is also supported by the participants' reports of their own strategies; the holistic strategy of comparing an item to an entire stored object was heavily used only by the Ordered All participants.

The inclusion of the Random All task also allows another account for the Ordered All practice effect to be rejected. Although the Ordered All task might be expected to demonstrate large response time improvements with practice simply because it had considerable room for improvement at first (response times on this task start at 1817 msec), the Random All task had equivalent room for improvement but did not show a comparable practice effect.

Two replications of Experiment 1 provided evidence concerning the boundary conditions on the unitization process that appears to have occurred during the Ordered All task. In the first replication, the size of the objects was increased from 4 cm to 10 cm. This size increase resulted in a greatly reduced practice effect for the Ordered All condition. The other three conditions were hardly affected. As such, there may well be a physical size constraint on unitization, such that only segments that can be viewed clearly without saccades can be unitized. In the second replication, the U-shaped figure below the five segments was removed, with the result that the stimuli were no longer closed figures. This alteration had little effect on the response times or practice effects for the four conditions. Thus, unitization can proceed even when the unit is a line rather than a closed object.

One possible objection to Experiment 1 is that the participants' subjective segments may not correspond to the five segments that were used to construct the stimuli. This is certainly possible; in fact, efforts were made so that it would be difficult to determine the experimental segmentation of an object simply by viewing it. The logic of the experiment does not assume that participants would naturally decompose objects into the five segments that were used to create them. The five segments themselves are fairly complex forms, and are likely to be composed out of segments themselves. The experiment logic simply requires that whatever composition process is required to identify a single segment, more composition is required to make a reliably accurate response in the All task. The assumption underlying the All task is that it cannot be reliably performed using a single functional unit that the participant had before the start of the experiment. The task either requires the separate identification of several components (that may themselves be built from smaller components), or the construction of a single component that did not exist prior to the experiment. Even if participants used an alternative parsing of the stimuli, for example by treating a part of A and a part of B as a single segment, then the five way conjunction still requires several of these segments under the plausible assumption that a single pre-existing segment will not cover all five experimenter-defined segments. Thus, for the present purposes it is not necessary to determine the actual segmentations participants originally used to break the whole objects into components.

In sum, Experiment 1 provides suggestive evidence that unitization of a complex stimulus occurs, but only when an image-like representation can be formed for it. The results are only suggestive because direct evidence has not yet been presented that response times are faster for the conjunctively defined object than would be predicted by an analytic model that integrates evidence from separately detected components. Unitization could produce responses that are faster than predicted by such an analytic model by requiring only a single component detection. This type of evidence will be considered in Experiment 5.

## Experiment 2

Experiment 2 explores boundary conditions on the unitization process suggested by Experiment 1. Specifically, Experiment 2 addressed the question of whether unitization requires stimulus components to be spatially connected. Whereas Experiment 1 manipulated unitizeability by presenting stimulus components in ordered versus random positions, Experiment 2 tests whether unitizeability is affected by presenting components in a contiguous versus separated fashion.

Previous research indicates that contiguity (spatial connectedness) plays a large role in perceptual parsing and organization. Palmer (1992, Palmer & Rock, 1994) has argued that one of the basic laws of perceptual organization is that stimulus parts tend to be grouped together if they are connected to one another. Baylis and Driver (1993) found that judging the relation between two points is facilitated if they come from the same object rather than separated objects.

Connectedness has also been found to influence concept learning more specifically. Shepp and Barrett (1991) found that children are better able to acquire a conjunctive categorization when the conjoined dimensions are connected together in the same object than when they are physically separated. Nahinsky et al. (1973) similarly found that physical proximity of stimulus parts facilitated acquiring concepts that involved several of the parts. Finally, Saiki and Hummel (1996) found that concepts based on the conjunction of the shape of an object and its spatial location relative to a second object were much more easily acquired when the two objects were physically connected than when they were separated. All three of these results indicate that conjunctions of stimulus parts are more efficiently used for category learning when the parts are contiguous. Given that Experiment 1 also involved the acquisition of a concept based on the conjunction of several parts, these previous studies might be interpreted as indicating that unitization should be greater for contiguous than separated components.

The prediction derived from the literature on contiguous and separated stimuli differs from the prediction made by the "imageability" hypothesis. According to the latter hypothesis, unitization can proceed as long as a single image-like representation can be formed for the set of components, and if the components are not separated by too great a retinal distance. In fact, some previous research has indicated that contiguity is not a strict constraint on unitization. Czerwinski et al. (1992) found unitization of sets of three disconnected line segments. Although disconnected, a single image could be formed for the set of line segments because of their consistent arrangement and small visual angle subtended. Similarly, Townsend, Hu, and Ashby (1981) found little influence of feature contiguity on the pattern of dependence they observed between detected features.

### Method

Participants. Seventy-six undergraduate students from Indiana University served as participants in order to fulfill a course requirement. The students were evenly split into the four between-subject conditions.

Materials. The design and construction of the "Contiguous" materials was identical to the ordered condition of Experiment 1, except that the lower bowl-shape was removed so as to provide a cleaner comparison to the Separated condition items. The "Separated" materials were constructed as in Experiment 1, except that the five segments within a stimulus were disconnected, and arranged vertically, as shown in Figure 4. The total longest distance between segments was equated in the separated and contiguous conditions. The distance from the left tip of the left-most segment to the right tip of the right-most segment of stimuli in the contiguous condition was equal to the distance from the top of the top-most doodle to the bottom of the bottom-most doodle in the separated condition.

Procedure. The procedural details closely followed the procedure used in Experiment 1, with the following exceptions. The participants were evenly divided into four groups created by factorially combining the two levels of stimulus appearance (Contiguous or Separated) with two levels of task type (All or One). In all cases, the stimuli were configured as they were in the Ordered condition of Experiment 1. That is, the Item ABCDE always had the same components A, B, C, D, and E in the same relative positions on each trial.

Results

The results of primary interest concern the correct response times for categorizing the Item ABCDE into Category 1. These response times are shown in Figure 5. A 2 (Separated vs Contiguous display) X 2 (All vs One component necessary) X 4 (blocks) ANOVA was applied to these data. A main effect of display was found, $F(1, 69) = 7.2$, mse = 14.4, $p < .01$, such that Contiguous displays were more quickly categorized than Separated displays. A main effect of task was found, $F(1, 69) = 18.2$, mse = 12.9, $p < .01$, with the One condition significantly faster than the All condition. Finally, a main effect of blocks was found, $F(3, 207) = 11.9$, mse = 168.2, $p < .01$, indicating that practice increased speed of categorization.

There was also a significant three-way interaction between these three variables, $F(3, 207) = 3.3$, mse = 9.5, $p < .05$. Although not obviously apparent in Figure 5, this interaction was due to a particularly strong practice effect in the All, Contiguous condition. The two All conditions became more separated with practice than did the two One conditions. The magnitude of this three-way interaction was several times smaller than it was in Experiment 1. If attention is restricted only to the two All conditions, a significant interaction between blocks and display was found, $F(3, 71) = 3.9$, mse = 11.2, $p < .05$. This interaction indicates that although the two All conditions appear roughly parallel after the first block of practice, the contiguous condition showed greater improvement with practice than did the separated condition.

Although cross-experiment comparisons are not strictly appropriate, it is notable that the practice effect for the Separated displays of Experiment 2 was much more pronounced than the practice effect for the Random displays of Experiment 1, $F(3, 207) = 6.8$, mse = 15.9, $p < .01$. As such, it appears that randomizing the order of components within a stimulus interfered with learning more than did separating the components but retaining their positions.

Power functions were fit to each participant's response times on each of the four blocks from each of the four conditions, yielding average best-fitting functions of

$$RT_{\text{All, Contiguous}} = 0.150 + 1.94\ N^{-0.528}$$

$$RT_{\text{All, Separated}} = 0.025 + 2.41\ N^{-0.32}$$

$$RT_{\text{One, Contiguous}} = 0.426 + 0.411\ N^{-1.93}$$

$$RT_{\text{One, Separated}} = 0.644 + 0.592\ N^{-5.67}$$

ANOVAs on the three constant values indicated no influence of condition on the best-fitting value for constant a (asymptote), $F(3,207) = 2.4$, mse = 0.55, but influences of condition on both b (total improvement), $F(3, 207) = 4.4$, $p < .01$, mse = 0.41, and c (learning rate), $F(3, 207) = 5.2$, $p < .01$, mse = 1.51. Even more strikingly than in Experiment 1, the conditions that were well fit by large b values tended to be well fit by small values of c. The Contiguous and Separated All conditions required much larger values of b than were required for their respective One conditions, with the Separated condition fit by a non-significantly larger value of b, and non-significantly smaller value of c, relative to the Contiguous condition. Using the magnitude of b as a measure of the improvement due to unitizeability, there is no evidence for greater unitization in the Contiguous than Separated condition.

In the All condition, response times for the other items were faster than for ABCDE. Overall, collapsing across the contiguous and separated groups, the Category 2 items were responded to in 1269 msec, which is significantly different from the 1387 msec required to respond to ABCDE, paired $t(75) = 8.2$, $p < .01$. Similarly, although ABCDE occurred more frequently than VWXYZ, the average correct response time for categorizing the latter object, 1196 msec, was significantly faster, paired $\underline{t}$ (75) = 6.4, $\underline{p} < .01$. Consistent with the unitization hypothesis, the speed advantage of Category 2 items over ABCDE in the All condition decreased over blocks of practice, as shown by a significant Item X Blocks interaction, $F(3, 103) = 8.0$, mse = 11.8, $\underline{p} < .01$. That is, over practice, ABCDE becomes relatively quickly categorized when compared to the other items.

The response time effects were mirrored by error analyses. Across the four blocks, the percentage of errors were 6.8%, 4.3%, 3.8%, and 3.7%, $F(3, 207) = 4.7$, mse = 1.2, $\underline{p}< .01$, indicating improvement from the beginning of the experiment to the end. The error rates for the four conditions were: Separated All = 5.6%, Separated One = 3.6%, Contiguous All = 5.9%, and Contiguous One = 3.5%, $F(3,207)$ = 3.8, mse = 1.0, $\underline{p} < .01$. These error rates exhibit the same ordering as the response times such that the slowest task was the most error prone. There was no significant interaction between condition and blocks on error rates, $F(9, 621) = 1.1$, mse = 1.5, $\underline{p} > .1$.

Discussion

Overall, the results indicated unitization for displays from the All condition that had separated components. The contiguous display showed greater improvement than did the separated display, using the response time difference across blocks as a measure. The effect size of this interaction was not large, but was statistically reliable. On the other hand, the separated condition showed much more improvement with practice than did the random order condition from Experiment 1, and the best-fitting power function indicated just as large a b value for the separated as for the contiguous All condition[1]. The clearest result from Experiment 2 is that physically separating components that must be integrated for a conjunctive judgment does not interfere with the unitization process to nearly as great extent as does randomly positioning these components.

The strong practice effects that were found in the separated condition support the notion that practice effects in a conjunctive task depend predominantly on establishing a consistent image-like representation for the unit to be conjoined. Consistent with Czerwinski et al's (1992) results, the current results suggest that unitization occurs despite the lack of physical contiguity. Unitization seems to depend strongly on the consistency with which a conjunctively defined item is rendered, at least in conditions where the physical separation between the ends of the stimuli is equated across the contiguous and separated conditions.

The separated condition is helpful in assessing explanations of the observed practice effects that do not posit unitization. One such explanation is that participants learn, with time, to attend to junctions between components. Given that the separated condition does not create diagnostic junctions between components, if identifying junctions were the main cause of practice effects then we would have expected large differences between the separated and contiguous conditions. Junction identification may still be occurring, and may explain the small but significant difference between the two display conditions. Still, robust practice effects do not depend on diagnostic junctions being found. As such, the results support unitization of single coherent images rather than the analytic integration of optimally diagnostic stimulus regions as the primary explanation for the strong practice effects that were observed.

One might explain the improvement in the separated condition in terms of learning what the different components look like and how they are associated with categories, rather than unitization. A problem with this account is that it would predict faster training in the separated than contiguous condition, given that the separated condition provides clearer information about what the individual components are. In fact, if anything, the steepness of the learning (the c parameter in the power law function) is greater for the

[1] Given that the difference between blocks' response times was greater for the Contiguous All than the Separated All condition, the fitted b value similarly might have been expected to be larger. This was not the case because the two All conditions did not achieve asymptotic levels by the fourth block, and the power function that was fit to the Separated All condition drops asymptotically below the Contiguous All condition's function.

Contiguous than the Separated condition. The comparable learning amounts (b parameters) and learning rates (c parameters) for these two conditions suggests that learning depends primarily on forming complex units rather than improving individual component registration times. Furthermore, by hypothesizing the same process of unitization for both conditions, an explanation is provided for their strikingly similar practice effects in terms of the amount and rate of learning.

## Experiment 3

In Experiment 2, the cohesiveness of the visual components to be integrated was manipulated by physically concatenating or separating them. In Experiment 3 a second method for manipulating the ease of integrating components was implemented. In this experiment, reliable categorization required attending to information from two experimenter-defined components. These components were always found within a five-segment curved contour. For one group, the two diagnostic components were next to and touching each other. For the other group, the two components were separated by a third component. Categorization times and practice effects were thus compared between situations where components that must be attended were together or apart.

The Together and Apart conditions in Experiment 3 controlled for different stimulus aspects than were controlled in Experiment 2. Experiment 2's separated and contiguous groups were equated for entire stimulus length, and maximal length between neighboring components. Experiment 3's groups were equated for curve contiguity but not length. As such, Experiment 3 tested whether the physical separation of components influences their unitizeability even when the components were joined together within a physically contiguous object. Two components were separated but part of a contiguous object by inserting a nondiagnostic component between them.

If the Together and Apart conditions differ, there are reasons to think that they may differ either by an absolute amount or that the degree of difference will be modulated by practice. The Together condition may show a persistent advantage across blocks of practice given that the Apart condition requires integration of information over a wider region. However, the Apart condition may also yield larger practice effects than the Together condition. One way of accomplishing the Apart task is to look for the conjunction of three contiguous segments - the two diagnostic components and the nondiagnostic component that lies between them. If this occurs, then greater improvement should be found in this task than in the Together task, under the assumption that unitizing a larger number of components produces a greater amount of practice. Of course, these hypotheses are not mutually exclusive; the Together task may continue to show an advantage over the Apart condition despite a greater influence of practice on the latter task.

### Method

Participants. Forty-eight undergraduate students from Indiana University served as participants in order to fulfill a course requirement.

Materials. The design and construction of the materials was highly similar to the previous experiments. Like Experiment 1, but unlike Experiment 2, the five components that comprised a curve were joined together by a U-shape to create a closed figure. Examples of the two conditions, Together and Apart, are shown in Figure 6. For both conditions, the single object belonging to Category 1 can be expressed as “ABCDE,” signifying that five components were in the stimulus, and that they were constrained so that each was unique, chosen randomly from a set of 10 components. Unlike the previous experiments, only this one item belonged in Category 1; there was no VWXYZ item in Category 1. Although removing VWXYZ from Category 1 means that each component has some diagnosticity for categorization, it has the benefit of creating a more natural and coherent Category 1 than was used in Experiments 1 and 2. Two objects belonged in Category 2, and each of these differed from ABCDE along a single component. The components along which the Category 1 items differed from the Category 2 item can be called “diagnostic” because they distinguished between the categories. In the Together condition, the two diagnostic components, of the Category 1 item were adjacent. In the Apart condition, the two diagnostic components of ABCDE were separated by a third nondiagnostic component that laid between them. The changes in the diagnosticity of the components of ABCDE in going from the Together to the Apart condition were made by changing the set of items in Category 2.

Figure 6 shows an example of the category items from the Together and Apart conditions. As shown in this figure, in the Apart condition, components D and E must be attended (or the junction between them) in order to reliably categorize ABCDE into Category 1. In the Together condition, components C and E must be attended. The labels "Together" and "Apart" thus refer to whether the two diagnostic components of ABCDE were adjacent to each other or not.

Figure 6 shows only one possible set of diagnostic components. Each participant in the Together condition was randomly assigned one of the four possible sets of adjacent segments to be the diagnostic segments, and each participant in the Apart together was randomly assigned one of the three sets of segments that were separated by one other segment: A and C, B and D, or C and E.

Procedure. The procedure closely followed that of the previous experiments. In all cases, the stimuli were configured as they were in the Ordered condition of Experiment 1. That is, the Item ABCDE always had the same components A, B, C, D, and E in the same relative positions on each trial. At the beginning of the experiment, participants were shown the objects that belonged to Category 1 and Category 2, and were asked to press a key when they had viewed the objects for a sufficient time to assure accurate categorization. Participants were given 320 trials in all. On each trial, one of the three objects was randomly selected (one in Category 1, and two in Category 2), and displayed in a random location on the screen. Participants' speed and accuracy at categorizing the object was recorded. The experiment required about 55 minutes.

## Results

As with Experiments 1 and 2, response time rather than accuracy was the more sensitive measure of participants' performance, and analyses focused on responses to the conjunctively defined object that belonged to Category 1. Figure 7 shows response times for correct categorizations of ABCDE. These results indicated a significant main effect of display type, $F(1, 47) = 18.4$, mse = 14.0, $p < .01$. The average response time in the Apart condition was 991 msec, compared to 757 msec in the Together condition.

A significant Block X Display Type interaction was also found, $F(3, 141) = 8.3$, mse = 10.4, $p < .01$, and is depicted in Figure 7. As shown in this figure, the Apart condition yielded a greater practice effect than did the Together condition. During the first block, the response time difference between the Apart and Together conditions was 290 msec. This difference was reduced to 165 by the final block.

Power functions were fit to each participant's response times on each of the four blocks, yielding average best-fitting functions of

$$RT_{Together} = 0.367 + 0.414\ N^{-1.230}$$

$$RT_{Apart} = 0.535 + 0.536\ N^{-1.841}$$

There were no significant differences between the Together and Apart conditions on the best-fitting values for a, unpaired $t(46) = 0.54$, b, $t(46) = 0.21$, or c, $t(46) = 0.68$. It may seem surprising that a higher value for c was fit for the Apart than Together condition, given that more prolonged, shallow, learning might be expected for the Apart condition under the assumption that it requires the formation of a more complex, 3-component unit. One possible reason for this is that the three-way conjunction can be constructed relatively quickly. Also, the practice curves are not at asymptotic level by the fourth block, and as such, there is some tradeoff in the power function between large b values and c values; large values of both b and c can accommodate large amounts of response time reduction (normally the domain of b alone) within the four pre-asymptotic blocks.

Response times were faster when the left-most or right-most component was diagnostic than when only middle components were diagnostic, $F(1, 47)= 4.5$, mse = 14.3, $p < .05$. If anything, this effect should have benefited the Apart condition more than the Together condition because 2 out of 3 of the configurations in the Apart condition had a diagnostic extreme component, whereas only 2 out of 4 of the Together configurations do.

The response time effects were directionally mirrored by error analyses but none were significant. Across the four blocks, the percentage of errors were 5.4%, 5.0%, 4.8%, and 4.4%, $F(3, 141) = 1.1$, mse = 1.7, $p > .1$. The error rates for the Together and Apart conditions were 4.7% and 5.1% respectively, $F(1, 47) = 0.7$, mse = 0.8, $p > .1$.

Discussion

A greater practice effect was found for the Apart than the Together condition, and this practice effect was not sufficient to equalize these tasks by the end of the experiment. That is, there was no evidence in the data that the practice effects were sufficiently large to eliminate the difference between display types, at least not after an hour of training. The block X condition interaction indicates that the total amount of response time improvement was greater for the Apart condition.

The primary result, that learning is greater for the Apart than Together task, is consistent with the hypothesis that participants approach the Apart task by developing units that span three components. While the Together task can be accomplished by conjoining two contiguous components, the Apart task must be accomplished by either 1) integrating evidence from two separated components, 2) creating a single unit that is not contiguous, or 3) creating a single contiguous unit that spans three contiguous components. The first possibility for the Apart task is eliminated by the pronounced practice effects that were observed. These practice effects are on the same order as those produced in Experiments 1 and 2 when unitization was possible, and are much larger than those found when responses required the integration of separate pieces of evidence. The second account for the Apart task is at odds with the results from Experiment 2. Experiment 2 did show that units could be formed from disconnected segments, however it also showed that the disconnected condition benefited less from practice than did the contiguous condition. Presumably, the attenuated practice effect was due to less coherent units being employed in the disconnected than the contiguous condition. In contrast to these results, in Experiment 3, the Apart condition showed greater practice effects than did the Together condition. Thus, participants in Experiment 3 were probably not treating the Apart condition by unitizing disconnected segments; if they had, weaker, not stronger, practice effects would have been predicted for this condition.

As such, the remaining hypothesis, that participants come to treat the Apart task by creating a three-component unit, seems to capture the current results and their relation to previous experiments most parsimoniously. By this account, the larger practice effect in the Apart than Together condition is due to the extended practice required to fuse a greater number of components together. The results are, in one sense, opposite to those produced in Experiment 2. In Experiment 2, the easier Contiguous task showed the greatest improvement with practice, while in Experiment 3, the easier Together task improved least. The comparison of Experiments 2 and 3 show that differential practice effects are not simply due to ceiling/floor effects. Easier tasks may be either more, or less strongly, affected by practice.

This account of Experiment 3 implies that people have a bias to form coherent, contiguous units. Experiment 2 shows that units can be formed when components are not contiguous, but Experiment 3 suggests that when components can be made contiguous by including extraneous, nondiagnostic information, then participants do so. This conclusion from Experiment 3 is consistent with work on attention suggesting that there is a strong bias for attention to be allocated to spatially contiguous regions (LaBerge & Brown, 1989; Eriksen & Murphy, 1987). It is typically difficult for people to attend to two separate regions without also attending the region in between (although Kramer and Hahn, 1995, find evidence for attention to discontiguous locations under some circumstances).

The experimental results do not imply that identifying a three-component unit will ever be as fast as is identifying a two-component unit. It might be argued that if a functional unit is really created, then it should not matter how many actual components are involved. This strong version of a unitization hypothesis is probably false. Just as all primitive features may not be equally quickly processed, constructed features (units) may also differ in how quickly they are processed, and it would be surprising if the actual physical space subtended by a unit did not affect its processing time.

## Experiment 4

Experiment 4 provides a parametric investigation of the relation between response time and the number of components that must be integrated to make a categorization. In essence, the intermediary cases between Experiment 1's All and One task were introduced, wherein participants must integrate information from 2, 3, or 4 out of the 5 components present.

The first goal of the experiment was to provide boundary conditions on the unitization process. Given components of a particular level of complexity, how many components can be unitized, and how long does it take to effectively unitize them? Although any results obtained will depend critically on the level of complexity of the primitive components, useful information can be obtained by observing whether linear changes in the number of segments to be combined give rise to linear changes in response time and practice effects. An upper limit in the number of segments that can be unitized should be revealed by a flattened practice effect. That is, if units that combine a certain number of segments cannot be created because they exceed a capacity limitation, then the practice effect across blocks should resemble conditions where unitization is difficult (e.g. Experiment 1, random order).

The second goal of the experiment is to explore the hypothesis, recruited to explain the difference between Together and Apart conditions in Experiment 3, that the time course of improvement for a conjunctive categorization task is positively related to the number of components that must be conjoined. By this hypothesis, the fewer components that need to be attended, the smaller the practice effects should be.

### Method

Participants. Eighty-five undergraduate students from Indiana University served as participants in order to fulfill a course requirement, and were equally divided into the five levels of the between-participants factor "number of components."

Materials. The materials were similar to those used in the previous experiments. All stimuli were composed out of five segments. Five conditions were included, in which the number of components that were required to make a reliably accurate Category 1 response were 1, 2, 3, 4, or 5. Thus, the "1" and "5" conditions were similar to the ordered One and All conditions from Experiment 1. The single item that belonged to category 1 can be represented as ABCDE for all five conditions; the item VWXYZ was not included. For the 1 condition, only a single item belonged in Category 2: ABCDZ. For the 2 condition, two items belonged in Category 2: ABCDZ and ABCYE. For the 3 condition, three items belonged in Category 2: ABCDZ, ABCYE, and ABXDE. For the 4 condition, four items belonged in Category 2: ABCDZ, ABCYE, ABWDE, and AVCDE. For the 5 condition, all five items shown under Category 2 in Figure 1 were used. Thus, unlike the previous experiments, the location of the diagnostic segments were not randomized. Whenever a segment was diagnostic for the item ABCDE of Category 1, all of the segments to the right of the segment were also diagnostic. In addition, the diagnostic segments always formed a contiguous curve, without any nondiagnostic intervening segments.

Procedure. As before, the participants' task was to categorize objects into Category 1 or Category 2 as quickly as possible, while maintaining categorization accuracy of at least 95%. Prior to the experimental trials, participants were presented with the objects that belonged to the two categories and their category memberships. Participants were presented with 400 trials in all, with equal numbers of Category 1 and Category 2 trials randomly intermixed.

### Results

Once again, response accuracies generally mirrored response times, but were less statistically sensitive to difference between conditions because of their restricted variability. For the primary analyses, only correct response times for the single Category

1 item were included. The response times for the five conditions, broken down by to block, are shown in Figure 8. A strong main effect of condition was found, $F(4, 81) = 132$, mse = 14.2, $p < .01$. Post-hoc tests revealed that all five conditions differed significantly from each other, $p < .01$. One particularly surprising result was that the 5 condition yielded significantly faster response times than did the 4 condition. Although this effect was not predicted, one explanation for it may be that participants in the 5 condition switched to the strategy of looking for the whole item "ABCDE" sooner than did participants in the 4 condition; from the instruction page itself, it would have been evident to these participants that all components were relevant for the categorization, and consequently they may not have initially attempted an analytic component-by-component integration. In addition, as Experiment 3 showed, participants were faster to respond when diagnostic components were positioned on the left- or right-most edges of the stimulus. Given that the left-most edge was diagnostic for the 5 condition but not the 4 condition, the diagnosticity of this highly salient component may have served as an additional impetus to create a unit that spanned the entire stimulus.

A strong main effect of block was also found, $F(3, 243) = 87$, mse = 9.5, $p < .01$, with all four of the blocks differing significantly from each other according to post-hoc tests. In addition to the two main effects, a significant interaction between block and condition was also found, $F(12, 243) = 8.3$, mse = 16.3, $p < .01$. As shown in Figure 8, this interaction can generally be characterized by a convergence of the five conditions over practice. There is a 946 msec difference between the conditions during the first block, and this reduces to a 490 msec difference at the final block. This difference cannot be explained away by using a ratio instead of an interval scale for evaluating practice effects. Over the course of practice, the 4 condition loses 38% of its initial value, while the 1 condition loses only 26% of its value.

Power functions were fit to each participant's response times on each of the four blocks, yielding best-fitting functions of

$$RT_{One} = 0.367 + 0.188\ N^{-2.080}$$

$$RT_{Two} = 0.490 + 0.379\ N^{-1.241}$$

$$RT_{Three} = 0.060 + 1.15\ N^{-0.438}$$

$$RT_{Four} = 0.02 + 3.07\ N^{-0.104}$$

$$RT_{Five} = 0.419 + 1.32\ N^{-.513}$$

There was no significant main effect of condition on the best-fitting value for constant a, $F(4,81) = 1.04$, mse = .465, $p > .05$. There were, however, significant main effects of condition on both b, $F(4, 81) = 8.4$, mse = 0.29, $p < .01$, and c, $F(4, 81) = 7.2$, mse = 0.23, $p < .01$. The influence of condition on b can be interpreted as showing that the total amount of learning increases as the number of components to be conjoined increases. This trend is strictly monotonic except between conditions Four and Five. The influence of condition on c indicates that the learning curves tends to become steeper (higher learning rate) as the number of components to be conjoined decreases. Thus, learning becomes more prolonged as the size of the conjunction increases, both in the sense of being less steep and more substantial. As a secondary analysis, the dependent variable "falling rate" was derived by finding response time differences between adjacent blocks. Submitting falling rate to a block X condition ANOVA revealed a significant interaction, $F(8, 162) = 7.1$, mse = 6.2, $p < .05$, indicating that the falling rate for many-component tasks was particularly large, relative to few-component tasks, as blocks increased.

Across the four blocks, the percentage of errors were 5.6%, 4.7%, 4.2%, and 4.3%, $F(3,243) = 2.7$, mse = 0.8, $p < .05$, indicating improvement from the beginning of the experiment to the end. The error rates for the five conditions were: One = 3.5%, Two = 4.0%, Three = 4.7%, Four = 5.7%, and Five = 5.7%, $F(4,81) = 3.8$, mse = 1.3, $p < .05$. There was no significant interaction between condition and blocks on error rates, $F(12,243) = 0.2$, mse = 1.0, $p > .1$.

Discussion

Experiment 4 confirmed the hypothesis that was introduced to explain the results of Experiment 3 -- as the number of segments that must be integrated increases, so does the magnitude of the improvement. This was tested parametrically in the experiment, and a nearly monotonic relation was found between the magnitude of the practice effect and the number of segments to be conjoined. The one unpredicted violation of this generalization was that the four segment task showed more facilitation with time, and longer initial response times, than did the five segment task. A post-hoc explanation for this violation may be that when a highly salient (e.g. see the results from Experiment 3) end segment is diagnostic for the conjunctive response, and all of the others segments are also diagnostic, participants have an early disposition to accomplish the task by matching each stimulus to their ABCDE image. This strategy, initiated early, may be more effective than a component-by-component integration strategy, which participants in the four-segment condition are more likely to adopt.

In addition to showing generally larger practice effects as the number of conjoined components increased, Experiment 4 also showed more protracted practice effects as the number of components increased, as indicated by the more rapid power function learning curves for tasks with few conjoined components. As such, learning is more gradual as the unitization task involves more components. This suggests one type of boundary condition/bottleneck on the unitization process. How quickly unitization can proceed depends on how much there is to unitize. This constraint argues against a radical version of the hypothesis that unitization simply involves creating a photograph-like image of the unit, and matching incoming stimuli to this image. Photographs do not take longer to develop if they involve more complex scenes, but the unitization process observed in Experiment 4 does depend on the complexity of the unit to be formed. A unit, once formed, may have image-like properties, but during its formation, it probably requires componential processing. The eventual image-like property that is observed in the units is the relative independence of categorization time from complexity, although complexity always exerts some influence. However, this end-state property seems to depend on a learning process that is quite sensitive to stimulus complexity.

It appears that there are boundary conditions on how much unitization can occur within a given period of time. The learning rates become progressively slower (smaller c values) as the number of components to combine for a conjunctive response increases. This limit on the number of components that can be integrated per trial is interesting because it stands in contrast to the lack of constraint that was observed in terms of the number of components that can be unitized. A failure of unitization would be shown by decreasing practice effects at some point, as number of components increased. Reduced practice effects were observed earlier when unitization was disrupted by randomly ordering components. The nearly monotonic trend for practice effects to increase as number of components increased suggests that if there is a limit to how many components can be unitized, it is greater than five. This result indicates an impressive unitization proficiency, given the complex nature of the components themselves.

## Experiment 5

The purpose of Experiment 5 was to compare an unitization account of categorization to specific analytic models. Thus far, large and gradual improvements in response times have been taken as evidence of unitization. This is also the type of evidence that has been generally used to argue for unitization in attention research (Czerwinski et al., 1992; LaBerge & Brown, 1989). However, it is possible to develop more specific analytic models of categorization, and test whether violations of these models are found. Thus, the two types of models that will be compared in Experiment 5 are a unitization model that assumes that a single functional unit is created for the set of five components composing a stimulus, and an analytic model that assumes that categorization decisions are based on the integration of five separate judgments about the components. The analytic model differs from the unitization account in that it does not assume the construction of functionally new features to subserve categorization. The analytic account can certainly predict improvements in responding to a conjunctively defined object, but these improvements would have to be due to improvements in integrating, identifying, or registering components.

## Deriving an Analytic Model

The general strategy in testing the independent-channels analytic model is to interpret the analytic model charitably, giving it several information processing advantages. Given these advantages, if categorization responses are still faster than predicted by the analytic model, then the analytic model would be discredited. The advantages that will be given to the analytic model are unlimited capacity processing, and parallelism (Townsend, 1990). Unlimited capacity means that the time required to identify one component is not slowed by the simultaneous requirement to look for another component. Parallelism means that any number of components can be processed simultaneously. Both assumptions serve to facilitate processing of the analytic model. Thus, even if they are unrealistic, the assumptions will err on the side of making the analytic model predict conjunctive categorizations that are too fast rather than too slow.

Intuition might suggest that an analytic model endowed with parallel, unlimited capacity processing would generate equal response times in the One and All tasks after practice. Surprisingly this is not the case, as is made evidence from two observations: A) there is natural variability in response times even in the One task, and B) the All task is intrinsically a conjunctive task. An example of how an analytic model can be derived is shown in Panels A and B of Figure 9. One starts by deriving the empirically obtained response time distribution for the One task. Then, five randomly sampled times from this distribution can be selected, and the maximum of these five times is determined, as indicated in the middle graphs of Panel A. The maximum, rather than the mean, is used because a response in the All task cannot be made until all five curves are detected in the object. The predicted response time distribution for the All task can be established by taking multiple random samples of five response times. Thus, there is a formal way of predicting the response time distribution in the conjunctive task, based on the response time distribution in the task that requires identification of only a single component. If this is done, then despite the charitable processing assumptions, this analytic model predicts response times in the conjunctive task to be slower than response times in the simple task, for the simple reason that the maximum of several samples of a random variable will typically be larger than the average of the samples.

Although it would be possible to derive the analytic models' prediction for the All task by randomly sampling sets of five times from the One task's distribution (as described above), fortunately there is an easier and more precise way to derive predictions, shown in Panel B. The analytic model's prediction for the All task can be found by computing the cumulative response time distribution for the One task, and then raising each value on this curve to the fifth power. The resulting distribution represents the analytic model's prediction for the All task, assuming that there are five components to identify. An example can provide the intuition behind this logic. If the probability of a response time being less than 300 msec. in the One task is .25, then the probability of five independently sampled response times from the same distribution being less than 300 msec is .00098 ($.25^5$). The conjunctive All task requires that all five components be identified before categorizing the Item ABCDE into Category 1. As such, the analytic model predicts that the reliably correct categorization of ABCDE should be made within 300 msec less than one time out of a thousand.

## Accommodating Dependencies in the Analytic Model

The psychological assumptions underlying the above algebraic treatment are: pure parallelism, unlimited capacity, and independently sampled response times. The first two assumptions are not problematic; they cannot be responsible for the analytic model predicting response times that are too slow because, if anything, they produce response times that are faster than they would otherwise be. However, the third assumption may be problematic, and additional experimental conditions and analyses are required in order to relax it.

It is first necessary to understand what the assumption of independent sampling means. According to the analytic model, five separate component identifications occur on every All task trial. According to independent sampling, the time required to identify one component is independent of the other identification times. Independence may be violated by either positive or negative dependencies. Negative dependencies would be expected if there were competition between components to be identified. If there were limited resources available for identifying components, then one component might be identified quickly if substantial resources were devoted to it, but this fast identification would come at the expense of the other components. Such negative contingencies are not problematic for the conclusions raised in Experiment 5. If negative contingencies occur then the algebraic formulation of the analytic model underpredicts actual analytic response times. If there are negative contingencies between sampled response times then the maximum of the five response times will be large relative to when there are no contingencies. Negative contingencies would yield even slower analytic predictions than those shown previously, entailing even more dramatic empirical violations of the analytic model.

However, if there is a positive correlation between identification times, then the analytic model could be faster than presented earlier. In fact, if there were perfect correlation between the sampled times, then the analytic model predicts that the All task would be performed as quickly as the One task. There are two important situations where positive contingencies are expected. Positive contingencies have been empirically observed when configural properties can be used instead of individual components (Townsend, Hu, & Kadlec, 1988). If this type of positive dependency is at play, then the conclusions are not in jeopardy. This type of positive dependency relies on an explanation very similar to the one presented; both argue for processing at a configural level above the individual component.

A second mechanism that produces positive dependencies is through shared input-output processes between the sampled response times. Imagine, for example, that on half of the One and All task trials, the response keyboard was moved ten feet away from participants. The One task's response time distribution would probably be distinctly bimodal, with half of the responses clustered around 1 second (the average detection time) and the other half clustered around 10 seconds (the average time to detect a component and walk toward the keyboard). In developing the analytic model's prediction for the All task, it would be inappropriate to sample five response times from the entire distribution. If this were done, then the maximum of the five times would be around 10 seconds about 97% of the time ($1 - (.5^5)$). This is inappropriate because 5 samples are taken from an underlying distribution that includes input and output processes as well as detection times. A more appropriate analytic model would sample 5 times from the distribution of the detection times and take the maximum, but then combine this time with a single sampled response time from the input and output distribution.

Fortunately, research on response times has provided a method for separating out various processes that make up a whole response time. The technique was introduced to psychology by Green and Luce (1971), and is described in detail by Smith (1990). The technique has been assessed, applied, and critiqued by (Sheu & Ratcliff, 1995). If a task requires two processes, A and B, and these processes each take a constant amount of time, and the time required for A is known, then to determine the time required for B one simply subtracts the time required for the A task by itself from the time required for the A+B task, assuming factor additivity. However, if the processes are stochastic, then each process will be associated with a distribution of response times rather than a constant. In this case, in order to determine the distribution of the unknown B process, the distribution of times in the A task can be deconvolved from the A+B task distribution. Just as the blur can be removed from a picture by deconvolving a Gaussian curve from the blurred image, so the influence of a particular process can be removed by deconvolving the distribution associated with the process from the whole distribution. The standard technique for this deconvolution is to calculate a Fourier transform for the whole and component distributions. The basic mathematical property of the Fourier transformation is that it converts convolution (or deconvolution) operation in the time domain into multiplication (or division) in the frequency domain. Using this property, the distribution associated with Task B can be determined as long as the A and A+B distributions are known, assuming factor additivity, by three steps: 1) take the Fourier transform of the A and A+B distributions, 2) divide the Fourier transform of the A+B distribution by the Fourier transform of the A distribution, and 3) take the inverse Fourier transform of the resulting quotient to derive the B distribution.

How is this technique used in eliminating certain types of positive dependencies between sampled response times? Positive dependencies due to shared input-output processes can be eliminated by obtaining an empirical distribution of input-output times for each participant. This distribution is obtained by having participants complete many trials of a simple detection task; the participant presses a specified key as soon as any curve is displayed. This task requires perceptual registration of the pattern and motor output, but does not require any decisions to be made based on the appearance of segments within the curve. This will be called the "simple detection" task. The general strategy, then, is to divide the One task into two processes: the input and output processes that should only be sampled once for each All task judgment, and the comparison/identification process which must be sampled five times according to the analytic model. To fully instantiate the analytic model, the following steps are taken: 1) Fourier transforms of the simple detection and One task distributions are calculated, 2) both Fourier transforms are filtered, 3) the Fourier transform of the One task is divided by the Fourier transform of the detection task, yielding the Fourier transform of the comparison distribution, 4) an inverse Fourier transform is calculated to convert the comparison distribution back to the time domain, 5) this new distribution is converted to a cumulative distribution, 6) every point on this cumulative distribution is raised to the fifth power to derive the analytic model's prediction, 7) the resulting distribution is convolved with the cumulative distribution for the simple detection task, and 8) the resulting cumulative distribution is compared to the empirically obtained All task distribution.

Final Considerations

The primary method of testing the analytic model will be to compare its prediction for the All task's RT distribution (derived from the One task) to the empirically observed distribution in the All task. A non-parametric Komolgorov-Smirnoff test for equal distributions can be used to see whether the analytic model is significantly violated.

In order to find violations of the charitably interpreted analytic model, far more practice will be needed than was used in the previous experiments. The formal analytic model was tested on the results from the previous experiments, and no violations in the direction of unitization were found. In fact, the analytic model predicted far faster response times for the All task than were obtained. However, this is not surprising, given the liberally interpreted (allowing for complete parallelism and unlimited capacity) analytic model tested. To provide a situation in which the RT distributions could possibly be faster than predicted by the analytic model, participants in Experiment 5 were given extended practice over more than 15 hours and 8000 trials.

The deconvolution analysis of response time distributions assumes the input/output distribution obtained from the simple detection task combines additively with the detection time distribution. One task strategy that would jeopardize this assumption would be if participants in the simple detection task learned to predict when to expect a stimulus to appear, and gave faster response times in this task than would be expected if they had to trigger their responses solely on the basis of the stimulus onset itself. For this reason, stimuli in all three tasks (One task, All task, and Simple Detection task) were presented after randomly determined temporal intervals.

Method

Participants. Four undergraduate research assistants from Indiana University were used as participants in this experiment. All four participants were naive with regard to the hypothesis being tested. The research assistants ran themselves in the experiment.

Materials. The materials were similar to those used in previous experiments. For the All tasks, one category contained ABCDE and the other category contained all five of the one-component distortions from ABCDE. In the One task, one category contained ABCDE and the other category contained only one of the five one-component distortions from ABCDE. For the simple detection task the only stimulus was ABCDE. Given the small number of participants used in the experiment, a different critical segment was assigned to each participant in the One task. As with the previous experiments, the actual instantiation of the components was randomized, under the constraint that none of the components used in the All task were used in the One task. To meet this constraint six new components were created. Once a random assignment of physical components to experimental letters was created, the assignment was retained for all of the sessions that a participant completed.

Procedure. The basic procedure from Experiment 1 was used. Participants categorized objects as quickly as possible. Unlike previous experiments, the type of task (One, All, and Simple Detection) was treated as a within-subject variable. Each session consisted of 500 trials, and lasted approximately 50, 45, and 35 minutes for the One, All, and Simple detection tasks respectively. Participants completed two sessions in one day, separated by a break of at least 10 minutes. The two sessions within one day were always of the same task type. Participants were instructed to alternate between the One and All tasks on successive days. All four participants completed 12 2-hour days of experimentation, consisting of 5 days of All task training, 5 days of One task training, and 2 days of simple detection task training. No significant influence of training was found for any participant during the simple detection task, and thus a single response time distribution was obtained for each participant that collapsed across all simple detection trials. The interval between days of experimentation varied, but the entire experiment was completed within 22 days for all participants.

for the All and One tasks, participants saw an equal number of Category 1 and Category 2 trials. In the All task, on trials in which a Category 2 item was to be selected, it was randomly selected from one of the five alternatives. Participants received trial-by-trial and block-by-block feedback on their categorization accuracy and response time. Participants were told to respond as quickly as possible while maintaining an error rate less than 5%.

For the Simple Detection task, at the beginning of each trial participants saw either a ‘1’ or ‘2’ displayed in the center of the screen for 300 msec, and they were instructed to press the key designated by the ‘1’ or ‘2’ regardless of the appearance of the stimulus. After a blank interval of variable length duration, the curve ABCDE was presented on the screen in a location randomized as it was in the previous experiments. It remained on the screen until a participant made a response, at which point the participant’s response was displayed on the screen for 300 msec. Whether ‘1’ or ‘2’ was the designated response was randomized. Two responses were used to make the motor control needed as similar as possible to that used in Experiment 5.

As opposed to the previous experiments, for all three tasks (All, One, and Simple Detection) the stimulus was displayed after a variable length delay so that participants could not predict when the stimulus would appear. Each trial was initiated by a warning signal, which was either “1” or “2” for the Simple Detection task, or a dot for the All and One tasks. As shown in the timeline for a single trial in Figure 10, after the warning signal was removed from the screen, a variable length delay occurred in which a blank screen was displayed for at least 300 msec. A geometric distribution was chosen for the variable length delay such that the probability of using a delay of X msec was equal to $(1 - .007)^{x-1}.007$. This distribution was used because it has the property that the probability of the delay ending at any given moment is independent of the time already elapsed.

## Results

Unless stated otherwise, all of the results include only correct Category 1 trials because these are the only trials that require a five-way conjunction in the All task. Figure 11 shows each of the four participants’ average response times across the sessions. The results from all four participants showed large practice effects, and a significant interaction between sessions and type of task, $F(3, 2997) > 25$, $p < .001$. Replicating the previous experiments, the All task showed much larger practice effects on RT than did the One task. One of the participants showed no systematic practice effect on the One task at all.

On the final session of training, a statistically reliable advantage for the One task over the All task was still found for all four participants, $F(1,999)>7.0$, $p < .01$. However, despite the faster average response times in the One task than in All task, the analytic model may still be violated. As a first step in testing the analytic model for the All task, the cumulative response time distributions for the final sessions of the All and One task were calculated for each participant. These cumulative distributions are shown in Figure 12. Inspection of Figure 12 reveals that for one of the participants (P.T.), the actual cumulative distribution for the All task was shifted to the left of the $One^5$ cumulative distribution. This dominance relation indicates a potential violation of the analytic model in that the All task was faster than predicted by the $One^5$ cumulative distribution. However, this analysis still assumes independent sampling of five response times from the One task distribution. For the other three participants, the cumulative distributions for the All task and the $One^5$ distribution cross at least once at an intermediary response time. J.W., S.P., P.T., and C.E. attained 95.1%, 94.4%, 94.2%, and 97.6% accuracy rates respectively on the final session of the experiment.

The average response times in the simple detection task for participants J.W., S.P., P.T., and C.E were 260, 295, 273, and 311 msec. respectively. The average standard deviations for these response times were 48, 49, 57, and 53 msec respectively. In general, the distributions had a slight positive skew.

These simple detection task times were collected to be deconvolved from the One task response time distribution, so as to provide an estimate of the identification response time distribution for one component. An example of the Fourier deconvolution technique is shown for participant J.W. in Figure 13. J.W.'s simple detection response time distribution is shown superimposed with his One task response time distribution. Fourier transformations were applied to each of these curves, and a parabolic filter with a cutoff of 20 was applied to each of the resulting frequency-domain representations. The transformed One task distribution was then divided by the transformed simple detection distribution, and an inverse Fourier transform was applied to the result to return to the original time domain. The result of these steps is shown by the curve labeled "deconvolution" in Figure 13. This curve represents the portion of the One task that is due to the actual comparison processes involved in deciding whether a particular curve segment provides evidence for a Category 1 response. If the comparison process distribution were convolved with the simple detection distribution, the original One task distribution would be reconstructed closely. The mild sinusoidal component present in the comparison process distribution is an artifact of the Fourier transformation process, and introduces only a small amount of error in the reconstruction.

For each of the participants, the One task distribution was deconvolved into simple detection and comparison distributions. With these separated distributions, the analytic model was implemented by raising each point on the cumulative comparison distribution to the fifth power, and convolving this resulting distribution with the cumulative simple task distribution. This resulting distribution represents the analytic model's predictions for the All task, assuming 1) five times are sampled independently from the comparison distribution and the maximum time is selected, 2) only one time is selected randomly from the simple detection task distribution, and 3) these two selected times are additively combined to make the prediction for the All task. The analytic model's predictions are shown in Figure 14, and the observed results from the All task are superimposed.

The results from all four participants indicated faster than predicted response times if we restrict our attention to response times that are faster than average. To test whether this violation is statistically reliable, a Komolgorov-Smirnov test was conducted (Amssey, 1951). This test finds the maximum discrepancy, D, between two cumulative distributions. The D statistic can be converted to an approximate chi-square variable by calculating

$$x^2 = 4D^2\left(\frac{N_1N_2}{N_1+N_2}\right),$$

with 2 degrees of freedom, where the $N_i$ is the sample size of distribution i. The Komolgorov-Smirnov tests were range-restricted to cumulative probabilities less than 0.5, as described by Birnbaum and Lientz (1972). This range restriction was selected because of previous work indicating that violations of an analytic model's predictions were found only for response times that were faster than average (Goldstone, 1997). With this restriction in place, the test statistic was significant at the $p < .01$ level for J.W., P.T., and C.E, but was only marginally significant for S.P at $p<.07$. As such, the fastest half of All task response times were significantly faster than predicted by the analytic model for three participants, and nonsignificantly faster for the fourth participant.

Comparisons of the All task's response time distribution to the analytic model were also made for the earlier sessions. For the first four sessions (two All task sessions and two One task sessions), the analytic model predicted far faster response times than were observed in the All task distribution for all four subjects, as measured by a Komolgorov-Smirnov test. The only reliable violations were found starting on the fifth sessions.

Discussion

The primary result from the experiment and modeling is that responses in the All task were faster than predicted by an unlimited capacity, purely parallel model that integrated responses from five separately registered components. Violations of this model were found assuming independence between the registration times for the five components, negative dependencies, and one class of positive dependencies -- namely, positive dependencies due to shared input-output processes. Despite the charitably interpreted analytic model, responses in the All task were reliably faster than predicted by the analytic model in certain circumstances. In particular, for the fastest half of the response times in the final sessions, the All task's cumulative response time distribution was reliably shifted to the left of the analytic model based on the corresponding One task's distribution. The two conditional restrictions will be discussed separately.

First, why were the violations of the analytic model restricted to the final sessions of training? By the unitization account, the analytic strategy would be the only one available to participants early in training, but as training continues participants would be expected to categorize according to the single unit with increasing frequency.

Second, why were the violations of the analytic model restricted to the fast response times? This range-dependent violation of the analytic model could potentially be accounted for by fast guesses in the All task, but this account is made less plausible by the overall high categorization accuracies. a more likely possibility is that a range of strategies was used for placing ABCDE into Category 1 in the All task. On some trials, an analytic strategy of combining evidence from separately detected components might have been used. On other trials, participants may have detected a single constructed unit. On trials where a participant used the analytic strategy, the charitably interpreted analytic model would be expected to underestimate observed response times, given the implausibility of pure parallel, unlimited capacity processing. However, on trials where participants used the single constructed unit to categorize ABCDE, violations of the analytic model are predicted. Participants would be expected to respond as soon as either the analytic or unit-based process produced a categorization (Logan, 1988). On average, the unit-based trials will be faster than the analytic trials. That is, if a participant successfully uses a single unit to categorize ABCDE then they will tend to do so quickly. If they cannot use this route, then their response time will tend to be slower. Thus, if the fast and slow response times tend to be based on single units and analytic integration respectively, then we would predict violations of the analytic model to be limited to, or more pronounced for, the fast response times.

If the above analysis is correct and the slowest response times were more likely to be produced by a process of independent detection of components, then the fastest response times offer the best opportunity for uncovering a unitization process that exceeds analytic predictions. In any case, the critical point is that empirical response time distributions that are faster than predicted by the analytic model have a very different evidential status from distributions that are slower than predicted. If response times are slower than predicted, it may be because participants are not using an analytic strategy, but it may also be that an analytic strategy is used in which one of the strong assumptions of parallelism and unlimited capacity is violated. Underestimations are consistent with participants adopting an analytic strategy of combining evidence from five separately detected components, but with limited capacity or imperfect parallelism.

However, if response times are faster than predicted, then the large class of analytic models that do not incorporate interactions between detected components is disconfirmed. As such, the justification for concentrating on situations where the analytic model predicts response times that are too long is that these situations allow relatively unambiguous conclusions to be drawn. Overestimations of All task response times by the analytic model are much more problematic for the analytic model than are underestimations. Overestimations suggest that a different process leads participants to categorize ABCDE quickly. This alternative process seems to involve the construction of a single functional unit that can be registered in a fashion that does not involve combining five independent component detections.

In applying formal models to tasks, assumptions must be made. At a broad level, the purpose of Experiment 5 was to eliminate the need for the assumption of independently sampled times. However, to eliminate this assumption, other assumptions were required. Sheu and Ratcliff (1995) have discussed the dependence of the Fourier deconvolution technique on factor additivity. Factor additivity may appear to be a strong assumption. For example, contrary to the assumption of factor additivity, increasing the difficulty of component registration may slow input-output times. However, when interactions between simple input-output and identification processes have been observed, they have not been very large (Dzhafarov & Cortese, 1996). Furthermore, to the extent that these interactions exist, they argue against a strict decomposition into input/output processes and identification processes, and without such a decomposition the analytic account has no explanation for the violations observed in Figure 12. Still, further work will be necessary to test analytic models without requiring factor additivity. Furthermore, only one variety of positive dependency was tested. The results from the All experiment can still be accommodated by an analytic model if the model is allowed to have positive dependencies between sampled times. In fact, Townsend et al. (1988) have observed positive dependencies in some perceptual judgment situations. If one curve component is detected more quickly because another component has already been detected, then positive dependencies can arise. However, at this point, the analytic model becomes similar to the unitization account. Both assume that segments are not registered independently.

**General Discussion**

The five reported experiments explored speed increases in a conjunctively defined categorization task. These speed increases were hypothesized to be due to the development of functional units over practice. Pronounced improvements in categorization were found when, and only when, unitization was possible and advantageous. Pronounced improvements were found in a categorization task that required all components of an object to be identified, but not when the task could be solved by attending to only a single component (Experiment 1). Greater improvements were found when the relevant components were joined together to form a coherent image rather than randomly ordered (Experiment 1). Larger improvements were found when the components were physically connected than when they were disconnected, although large improvements were found in both cases, consistent with the hypothesis that unitization depends critically on being able to form a single image-like representation (Experiment 2). Although Experiment 2 showed that units can be developed that span disconnected components, Experiment 3 showed that participants have a bias to create contiguous units if possible. When a category was defined by two diagnostic components with a third nondiagnostic component in between them, the relatively protracted course of improvement suggested that participants were creating units that incorporated all three components. This experiment also showed that unitization effects were not well explained by identifying junction points between diagnostic components.

The amount and gradualness of improvement observed in a conjunctive categorization task was positively related to the number of components that needed to be attended (Experiment 4). Although this experiment did not find evidence for a capacity limit in the number of components that could be unitized (in the range from 1-5), it did find evidence for a boundary condition on the unitization process itself. In particular, the experiment suggested that there was a limit on the number of components that could be unitized within a given number of trials. As the conjunction of more components was required, the learning rate decreased but the total amount of learning increased. Finally, Experiment 5 explored response times on a conjunctively defined categorization task following 16 hours of categorization training. The distribution of response times was compared to an analytic model derived from the distribution of times from a task requiring participants to attend to only one component. Responses were faster than predicted by this analytic model, particularly for the fastest half of responses. The violations of the analytic model were consistent with a unitization account that based categorizations on a single comparison to a developed unit rather than the integration of evidence from independently detected components.

By fitting power functions to the separate experimental conditions, the practice effects accompanying unitization were decomposed into effects of speed and extent of learning. When significant, the parameter differences indicated that unitization was associated with a greater total amount of learning (b) and a more gradual learning curve (c). In Experiments 1, 2, and 4, gradual learning was correlated with a greater total difference between initial and final performance. This correlation is not consistent with the postulation of a single factor, strength of practice, that controls both the steepness and extent of practice. However, this correlation is consistent with a unitization account whereby conditions that afford unitization produce substantial improvement with practice but only after prolonged training. The working assumption has been that extensive learning is an indicator of unitization, but the empirical results indicate that gradualness of learning may also be a useful indicator. Complex units apparently require time to build.

The conclusions with respect to unitization do not depend on participants agreeing with the experimenter on what the underlying components are. Participants may well begin processing the curves into components smaller than the level of a single experiment-defined component. In fact, the components were deliberately chosen to be complex units themselves. However, the comparison of the single-component One and multi-component All tasks in Experiments 1-3 is still appropriate. The One tasks may require the conjoining of image elements, but the All tasks require these conjunctions plus additional conjunctions, and it is the process of unitizing these larger elements that the experiments target. The general finding that performance in the One task quickly reached asymptotic levels suggests that if it does require unitization, this unitization is relatively fast compared to the unitization for tasks that require the registration of image elements that span across several experimenter-defined components.

The structure of the categories in Experiments 1 and 2 may seem unnatural, given that the completely dissimilar objects ABCDE and VWXYZ belonged to one category while all of the nearest neighbors to ABCDE belonged to the other category. Such an “exclusive-or” category structure is probably not common, but was useful in assuring that the individual components had no diagnosticity for categorization by themselves. In addition, in Experiments 3-5, the VWXYZ item was eliminated, establishing a more natural category in which a critical level of similarity to the item ABCDE was required to be in Category 1 and every other item belonged to a second “miscellaneous” category.

## Constraints on Unitization

The five experiments offer conclusions about the constraints and boundary conditions on the unitization process. Most clearly, from Experiment 1, pronounced improvements in speed were not found when the five components to be noticed were randomly ordered within an object from trial to trial. This suggests that whatever process is getting faster in the well ordered condition, the same process is not at work (to the same degree) in the randomly ordered condition. For example, the speed up in the well ordered condition is probably not due to speed ups in the time required to identify individual segments, to localize segments, or to integrate evidence from separate segments. These processes are all equally required in the ordered and random conditions, or are required to a greater extent in the random condition. The primary candidate for a process that is involved in the ordered, but not random, displays is the construction of an unitary image for the conjunction of five components. The difference in speed up in the two conditions is consistent with unitization processes being restricted to situations where a single image can be formed.

The results with respect to image coherency are somewhat mixed. Experiment 2 showed large improvements for a conjunction defined by disconnected segments. This result argues against the speed ups observed in other experiments simply being due to increased efficiency at locating junctions between components. The disconnected stimuli have no junctions but still yield large improvements. These results are also consistent with Czerwinski et al's (1992) results suggesting unitization occurs in a conjunctive feature search task even when disconnected line segments were used as features. On the other hand, in Experiment 2 the connected condition led to greater improvements than did the disconnected condition, and the results from Experiment 3 indicated that participants created units that included irrelevant intervening segments. A reconciliation between these sets of findings is that the most important constraint on unit development is that they be represented by a single image. Only secondarily does the contiguity of the unit affect its creation.

A dynamic constraint on unit formation was suggested by Experiment 4 -- as the number of components that needed to be combined for a conjunctive categorization increased, so did the gradualness of improvement as tokened by parameter c of the power function. Thus, higher-order conjunctions were categorized more slowly, but their rate of improvement was relatively protracted as well. This constraint indicates a way in which unit formation is not equivalent to the development of a photograph-like image. Essentially, unit formation is sensitive to the complexity of the unit. If unit formation were simply a matter of developing a photograph, then complexity would have little influence. Once formed, units may be relatively unaffected by complexity (Experiment 5), but the formation of the units is adversely affected by increasing the number of components to be combined. In short, unit formation seems to require greater componential processing than does unit deployment.

Interpreting large improvements in a conjunctive task as evidence for unitization, Experiment 4 did not find capacity limits on unitization. This may simply be due to testing too narrow a range of complexities. Capacity limits may exist, but may only be found with more than five components of the complexity used in experiments. Still, the result is reminiscent of findings suggesting indefinitely large short term memories as long as chunks can be formed (Ericcson, Chase, & Faloon, 1980).

Mechanisms for Improvement in a Conjunctive Task

Experiment 5 was able to successfully eliminate some broad classes of analytic models of categorization following prolonged practice. In particular, analytic models that integrated evidence from five separately detected components to make a conjunctive response were shown to predict responses times that were slower than those obtained for three out of four participants (the fourth showed non-significant violations in the same direction). These violations were found only after several hours of practice, and were are restricted to the fastest response times. Despite the impressive violations, there are still two qualitatively different mechanisms that could account for the pronounced speed up of the conjunctive categorization: a genuinely holistic matching process to a constructed unit, or an analytic model that assumes interactive facilitation among the component detectors.

According to a holistic match process, a conjunctive categorization is made by comparing the image of the presented item to an image that has been stored over prolonged practice. The stored image may have parts, but either these parts are arbitrarily small or do not play a functional role in the recognition of the image. There is considerable evidence supporting the gradual development of configural features. Neurophysiological findings suggest that some individual neurons can detect conjunctions of features (Perett & Oram, 1993), and that prolonged training can produce neurons that respond to trained configural patterns (Logothetis, Pauls, & Poggio, 1995). Evidence from the recognition of faces (Tanaka & Gauthier, 1997), objects following prolonged training (Gauthier & Tarr,1997), trained letter-like stimuli (Shiffrin & Lightfoot, 1997), and words (LaBerge & Samuels, 1974) supports the idea that entire configurations are learned with practice, and that detection of whole configurations is dissociable from detection of parts of the configurations.

However, despite the impressive improvements over time observed in Experiment 5, the results are also compatible with analytic models that assume interactions between the components (Mordkoff & Egeth, 1993; Townsend et al., 1988). For example, if detecting one component of "ABCDE" facilitates detection of other components, then positive dependencies among the detection times can result.. Experiment 5 tests an analytic model that permits positive dependencies, and still finds violations of this model. However, in the latter model, the only positive dependencies allowed were those due to shared input-processes between sampled components, rather than mutual facilitations. Although the experiments cannot rule out all analytic models, the experiments do place general constraints on models of conjunctive categorization following extended practice. Models that integrate independently registered components can be rejected. Well practiced categorizations either depend on strong mutual facilitations between component processors, or do not depend on component processing at all. In either case, the process is appropriately labeled "unitization" in that the percepts associated with different components are closely coupled together (see Townsend, Hu, & Kadlec, 1988). In fact, an interactive facilitation mechanism could be seen as the mechanism that implements holistic unit detection at a higher functional level of description.

**References**

Allen, P. A., Wallace, B., & Weber, T. A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. Journal of Experimental Psychology: Human Perception and Performance, 21, 914-934.

Amssey, F. J. (1951). The Komolgorov-Smirnov test for goodness of fit. Journal of the American Statistical Association, 46, 405-409.

Baylis, G. C., & Driver, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. Journal of Experimental Psychology: Human Perception and Performance, 19, 451-470.

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: Evidence from normal participants and a computational model. Journal of Experimental Psychology: Human Perception and Performance, 24, 1011-1036.

Besner, D., & Johnston, J. C. (1989). Reading and the mental lexicon: On the uptake of visual information. in W. Marslen-Wilson (Ed), Lexical representation and process. Cambridge, MA: MIT Press. (pp. 291-316).

Birnbaum, Z. W., & Lientz, B. P. (1972). Tables of critical values of some Renyi type statistics for finite sample sizes. American Statistical Association Journal, 42, 870-877.

Bruner, J. A., & Postman, L. (1949). Perception, conception, and behavior. Journal of Personality, 18, 14-31.

Cattell, J.M. (1886). The time it takes to see and name objects. Mind,11,63-65.

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

Diamond, R., & Carey. S. (1986). Why faces are and are not special: An effect of expertise. Journal of Experimental Psychology: General, 115, 107-117.

Dzhafarov, E. N., & Cortese, J. M. (1996). Empirical recovery of response-time decomposition rules 1: Sample-level decomposition tests. Journal of Mathematical Psychology, 40, 185-202.

Ericsson, K., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. Science, 208, 1181-1182.

Eriksen, C. W., & Murphy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. Perception & Psychophysics, 14, 255-260.

Gauthier, I. & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. Vision Research, 37, 1673-1682.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes, 38, 2401-2428.

Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. Journal of Memory and Language, 27, 166-195.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General, 123, 178-200.

Goldstone, R. L. (1997). Categorization and unitization. Indiana University Cognitive Science Technical Report #186. Bloomington, IN.

Goldstone, R. L. (1998). Perceptual Learning. Annual Review of Psychology, 49, 585-612.

Goldstone, R. L., & Schyns, P. (1994). Learning new features of representation. Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society. (pp. 974-978). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Green, D. M., & Luce, R. D. (1971). Detection of auditory signals presented at random times: III. Perception and Psychophysics, 9, 257-268.

Gumperz, J. J., & Levinson, S. C. (1991). Rethinking linguistic relativity. Current Anthropology, 32, 613-623.

Harnad, S. (1987). Categorical perception. Cambridge University Press: Cambridge.

Harnad, S., Hanson, S. J., & Lubin, J. (1994). Learned categorical perception in neural nets: Implications for symbol grounding. in V. Honavar & L. Uhr (Eds.) Artificial intelligence and neural networks: Steps toward principled integration. Academic Press: Boston. (pp 191-206).

Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 16, 321-338.

Hebb, D. O. (1949). The organization of behavior. New York: Wiley.

Hubel, D. H., & Wiesel (1968). Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology, 195, 215-243.

Johnson, N. F., (1975). On the function of letters in word identification: Some data and a preliminary model, Journal of Verbal Learning and Verbal Behavior, 14, 17-29.

Johnson, N. F., Turner-Lyga, M., & Pettegrew, B. S. (1986). Part/whole relationships in the processing of small visual patterns. Memory & Cognition, 14, 5-16.

Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: O as letter or as digit. Perception and Psychophysics, 12, 457-460.

Julesz, B. (1981). Textons, the elements of texture perception, and their interaction. Nature, 290, 91-97.

Koffka, K. (1935). Principles of Gestalt psychology. New York: Harcourt Brace.

Kramer, A. F., & Hahn, S. (1995). Splitting the beam: Distribution of attention over noncontiguous regions of the visual-field. Psychological Science, 6, 381-386.

Kurtz, K. J. (1996). Category-based similarity. In G. W. Cottrell (Ed.) Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, 290.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. Psychological Review, 99, 22-44.

LaBerge, D. (1973). Attention and the measurement of perceptual learning. Memory & Cognition, 1, 268-276.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. Psychological Review, 96, 101-124.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Lane, H. (1965). The motor theory of speech perception: A critical review. Psychological Review, 72, 275-309.

Lin, E. L., & Murphy, G. L. (1997). Effects of background knowledge on object categorization and part detection. Journal of Experimental Psychology: Human Perception and Performance, 23, 1153-1169.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 3, 732-753.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning-curves : a test of the instance theory of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 883-914.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. Current Biology, 5, 552-563.

Mordkoff, J. T., Egeth, H. E. (1993). Response time and accuracy revisited: Converging support for the interactive race model. Journal of Experimental Psychology: Human Perception and Performance, 19, 981-991.

Nahinsky, I. D., Slaymaker, F. L., Aamiry, A., & O'Brien, C. J. (1973). The concreteness of attributes in concept learning strategies. Memory & Cognition, 1, 307-318.

Niedenthal, P. M., & Kitayama, S. (1994). The heart's eye. San Diego: Acaedmic Press.

O'hara, W. (1980). Evidence in support of word unitization. Perception and Psychophysics, 27, 390-402.

Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. Journal of Experimental Psychology: Human Perception and Performance, 10, 413-428.

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

Palmer, S. E. (1992). Common region: A new principle of perceptual grouping. Cognitive Psychology, 24, 436-447.

Palmer, S. E., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. Psychonomic Bulletin & Review

Perett, D. I., & Oram, M. W. (1993). Neurphysiology of shape processing. Image and Vision Computing, 11, 317-333.

Pevtzow, R., & Goldstone, R. L. (1994). Categorization and the parsing of objects. Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society. (pp. 717-722). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Rosenbloom, P. S., & Newell, A. (1987). An integrated computational model of stimulus-response compatibility and practice. Psychology of Learning and Motivation, 21, 1-52.

Saiki, J., & Hummel, J. E. (1996). Attribute conjunctions and the part configuration advantage in object category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 1002-1119.

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

Schyns, P., Goldstone, R. L., & Thibaut, J-P. (1995). The development of features in object concepts. Indiana University Technical Report #106. Bloomington, Indiana.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In Medin (Ed.). The Psychology of Learning and Motivation, 31, 305-354. Academic Press: San Diego, CA.

Schyns, P. G., & Murphy, G. L., (1993). The ontogeny of transformable part representations in object concepts. In Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, (pp. 197-202). Hillsdale, NJ: Erlbaum.

Schyns, P. G., & Rodet (1997). Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 681-696.

Shepp, B. E., & Barrett, S. E. (1991). The development of perceived structure and attention: Evidence from divided and selective attention tasks. Journal of Experimental Child Psychology, 51, 434-458.

Sheu, C-F, & Ratcliff, R. (1995). The application of fourier deconvolution to reaction time data: A cautionary note. Psychological Bulletin, 118, 234-251.

Shiffrin, R. M., & Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.) The Psychology of Learning and Motivation, Volume 36. San Diego: Academic Press. (pp. 45-82).

Smith, E. E., & Haviland, S. E. (1972). Why words are perceived more accurately than nonwords: Inference versus unitization. Journal of Experimental Psychology, 92, 59-64.

Smith, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time data. Psychological Bulletin, 108, 533-550.

Tao, L., Healy, A. F., & Bourne, L. E. (1997). Unitization in second-language learning: Evidence from letter detection. American Journal of Psychology, 110, 385-395.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology, 46A, 225-245.

Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.) The Psychology of Learning and Motivation, Volume 36. San Diego: Academic Press. (pp. 83-126).

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. Psychological Science, 1, 46-54.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight-line features. Psychological Research, 43, 259-275.

Townsend, J. T., Hu, G. G., & Kadlec, H. (1988). Feature sensitivity, bias, and interdependencies as a function of energy and payoffs. Perception and Psychophysics, 43, 575-591.

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

Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. American Journal of Psychology, 105, 341-362.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. Perception & Psychophysics, 56, 495-500.

Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object-superiority effect. Science, 186, 752-755.

Whorf, B. L. (1941). Languages and logic. in J. B. Carroll (ed.) Language, Thought, and Reality: Selected papers of Benjamin Lee Whorf. MIT Press (1956), Cambridge, Mass. (pp. 233-245).

Yin, R. K. (1969). Looking at upside-down faces. Journal of Experimental Psychology, 81, 141-145.

**Author Notes**

Many useful comments and suggestions were provided by Gordan Logan, John Kruschke, Douglas Medin, Robert Nosofsky, Roger Ratcliff, Philippe Schyns, Linda Smith, and two anonymous reviewers. Special thanks go to Ching-fan Sheu for providing code for Fourier deconvolution of response time distributions, and to Tom Busey, Stevan Harnad, Richard Shiffrin, and James Townsend for extended discussions and comments . I thank Chris Howard, Nicole Turcotte, Amy Hess, and Dan Manco for their assistance in conducting the experiments. This research was funded by National Science Foundation Grant SBR-9409232, a James McKeen Cattell award, and a Gill fellowship. Correspondence concerning this article should be addressed to rgoldsto@indiana.edu or Robert Goldstone, Psychology Department, Indiana University, Bloomington, Indiana 47405. Further information about the laboratory can be found at http://cognitrn.psych.indiana.edu/.

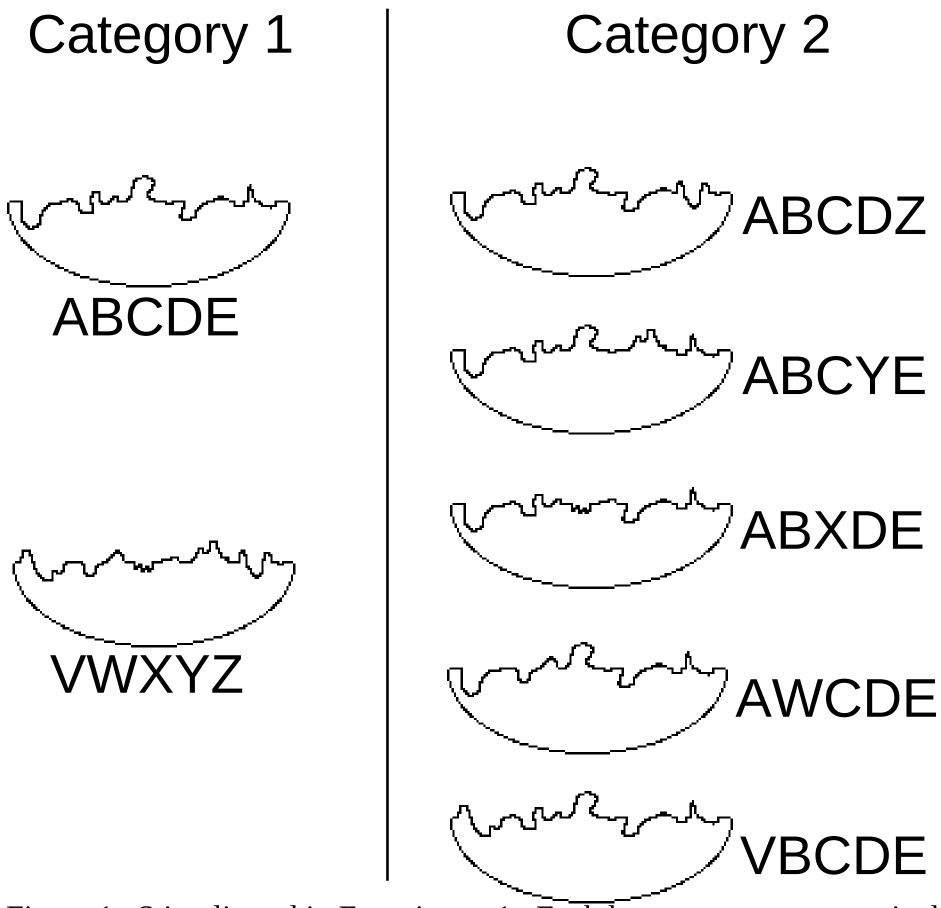


Figure 1. Stimuli used in Experiment 1. Each letter represents a particular stimulus segment, and each stimulus is composed of five segments. To categorize the item represented by “ABCDE” as belonging to Category 1, it is necessary to process information associated with each of the segments.

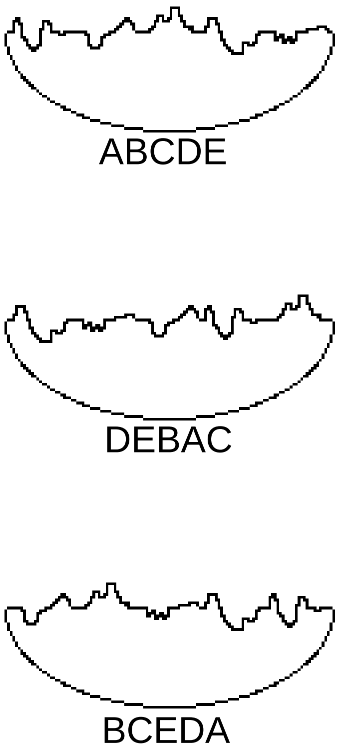


Figure 2. In the Random condition, any ordering of the same components counts as the same object. Thus, when object "ABCDE" is presented, it can be presented in 120 different ways, 3 of which are shown above. In the Ordered condition, the spatial positions of the five segments are fixed.

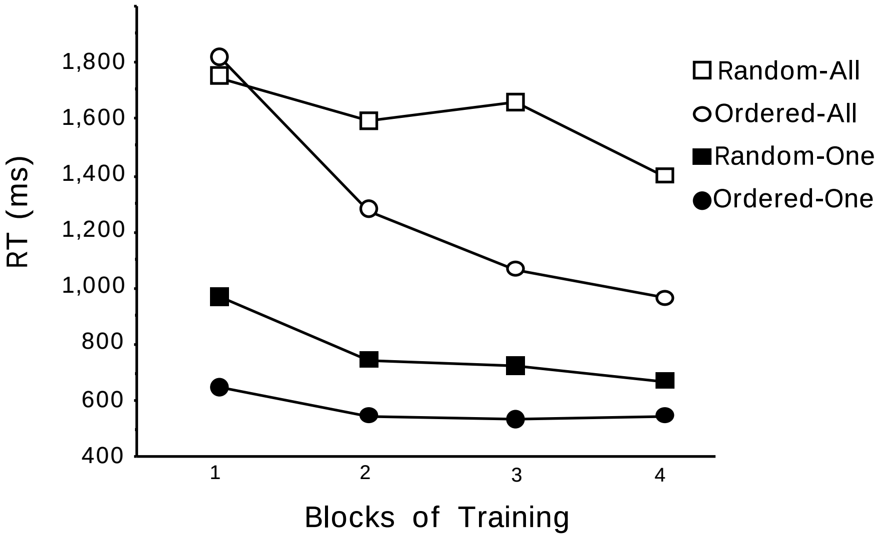


Figure 3. Results from Experiment 1. The most pronounced practice effects were observed for the conjunctive All task with constantly ordered components.

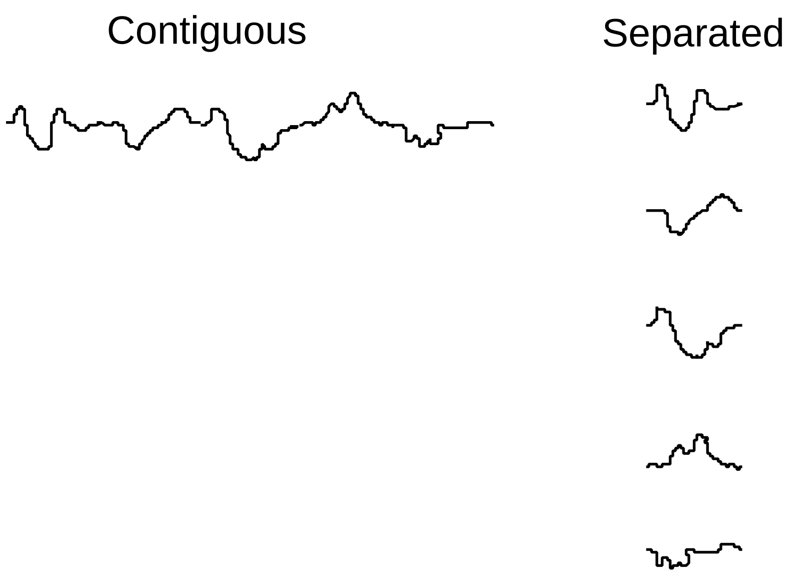


Figure 4. Contiguous and Separated stimuli from Experiment 2.

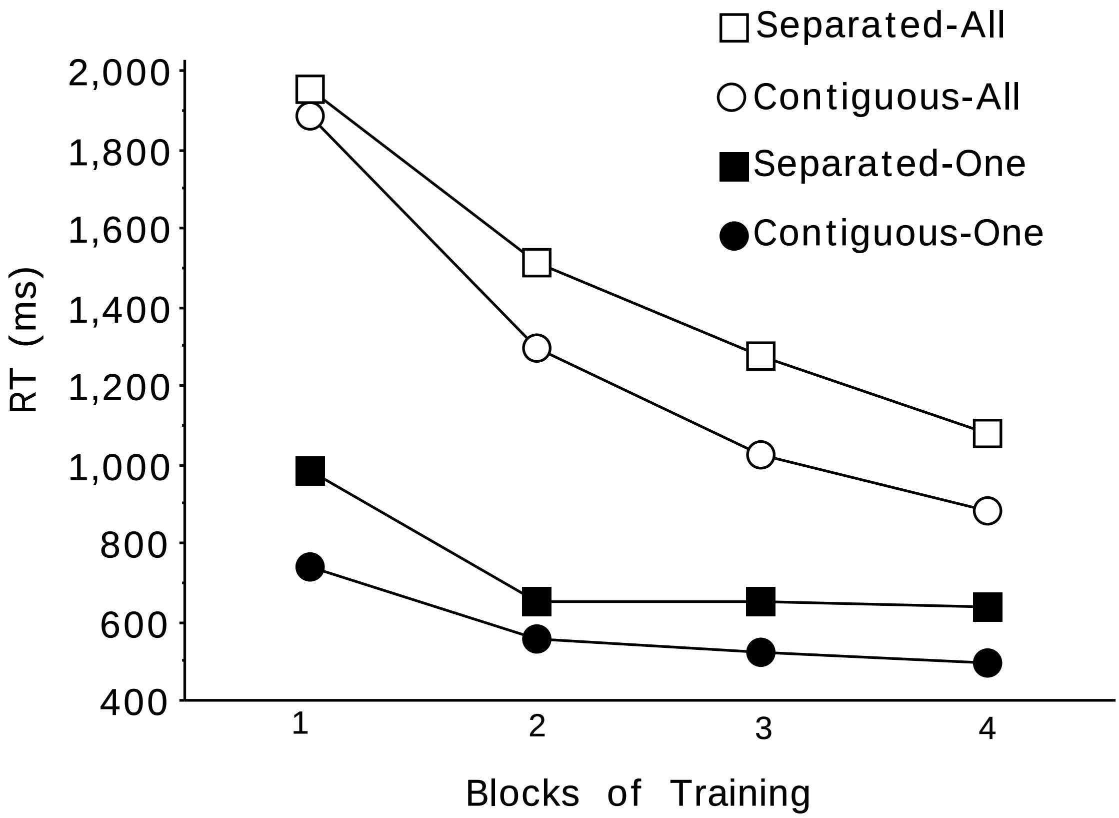


Figure 5. Results from Experiment 2. Although both conjunctive All tasks show large practice effects, the practice effects are slightly larger for the Contiguous displays.

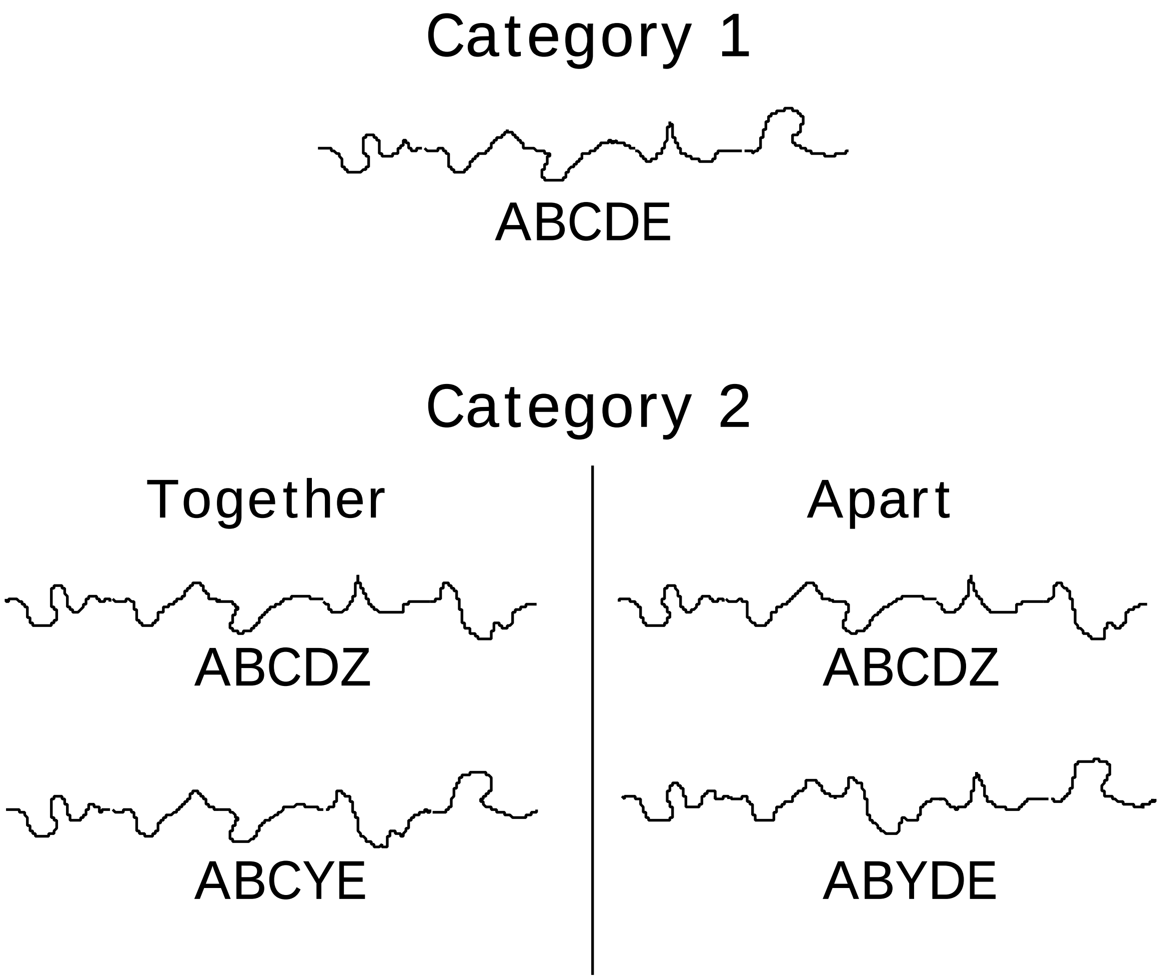


Figure 6. Stimuli from the Together and Apart conditions of Experiment 3. In both conditions, two segments were relevant for categorization. In the Together conditions, these segments were adjacent. In the Apart condition, they were separated by a third, nondiagnostic segment.

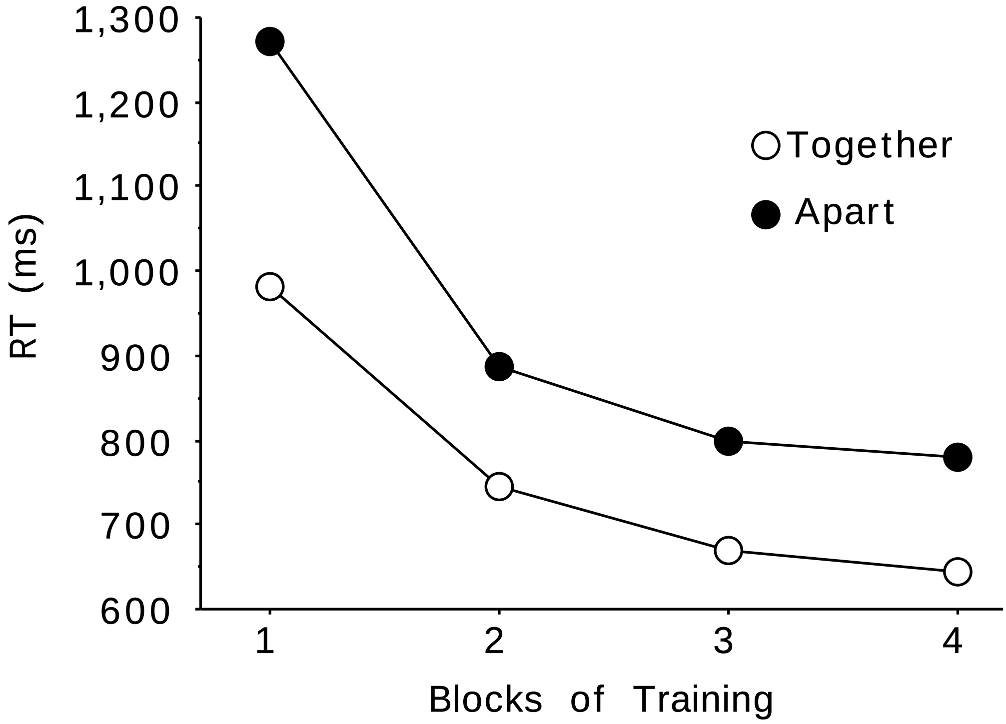


Figure 7. Results from Experiment 3. Larger and more protracted practice effects were observed in the Apart than in the Together condition.

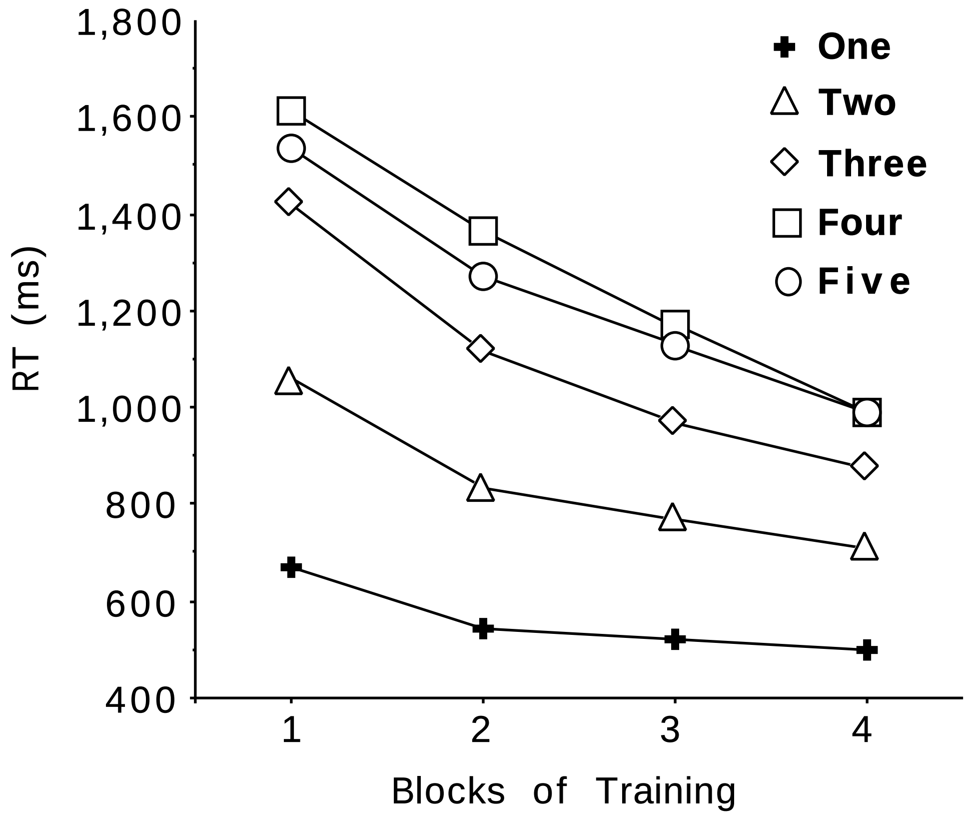


Figure 8. Results from Experiment 4. With the exception of the 4-way and 5-way conjunctions, practice effects were larger and more protracted as more components were required for the conjunctive categorization.

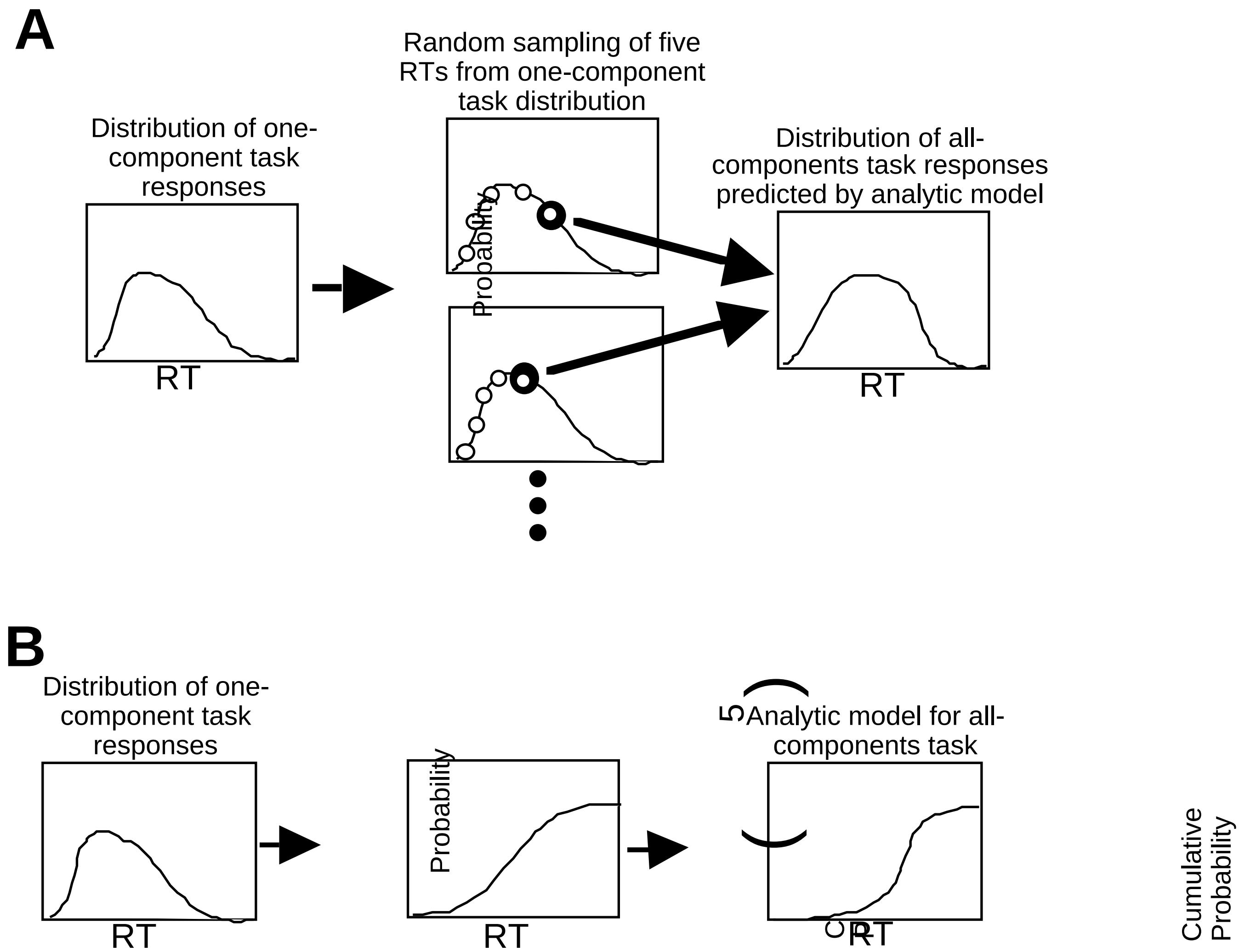


Figure 9. This figure depicts the initial procedure for deriving an analytic model's predictions for an All task from a RT distribution obtained from a One task. In Panel A, five RTs are randomly sampled, and a predicted distribution for the All task is formed by selecting the maximum of these times. Assuming independently sampled times, a simpler technique for creating the predicted All task response time distribution, shown in Figure B, is to raise every point of the observed One task's cumulative probability distribution to the fifth power.

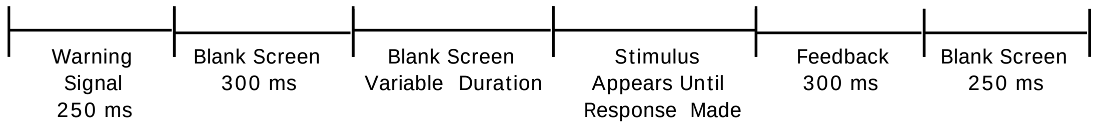


Figure 10. The time line for each trial in Experiment 4.

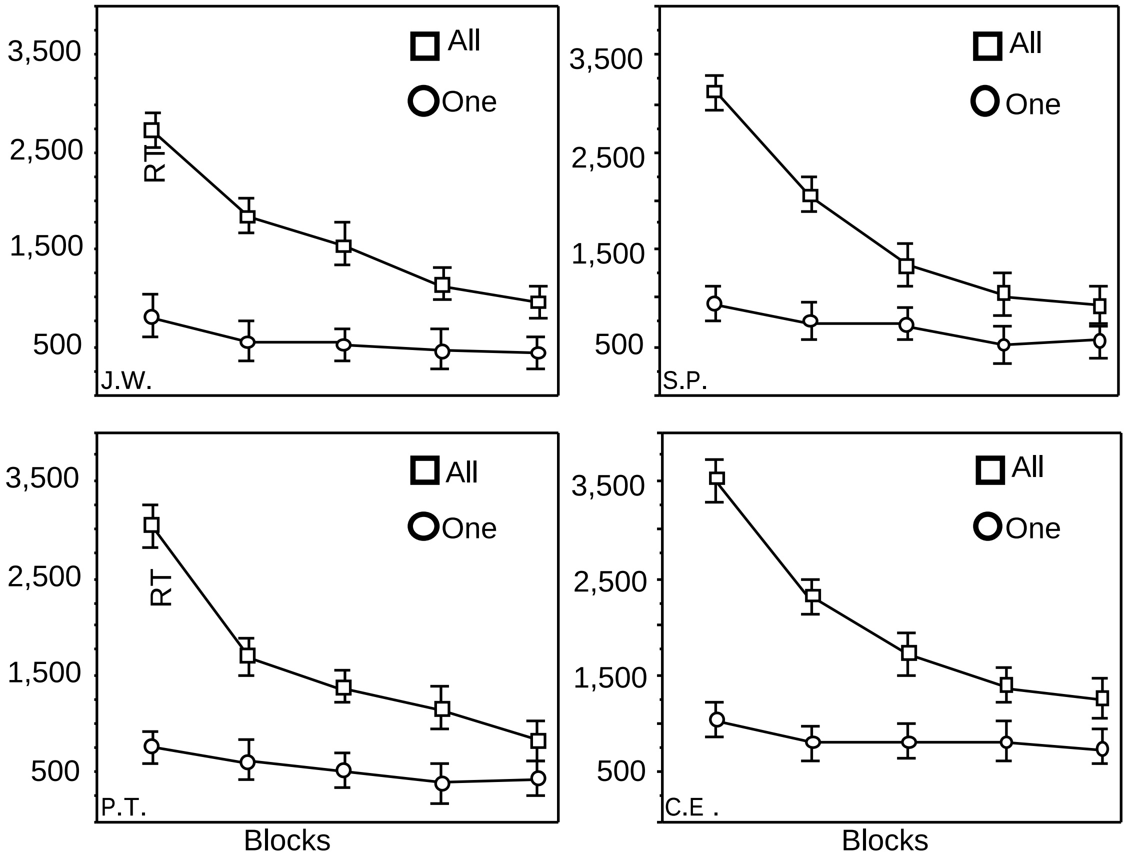


Figure 11. Results from Experiment 5, for each of the four participants. As with earlier experiments, the All task shows a much greater practice effect than does the One task. Practice effects in the All task were observed throughout the 16 hours of training.

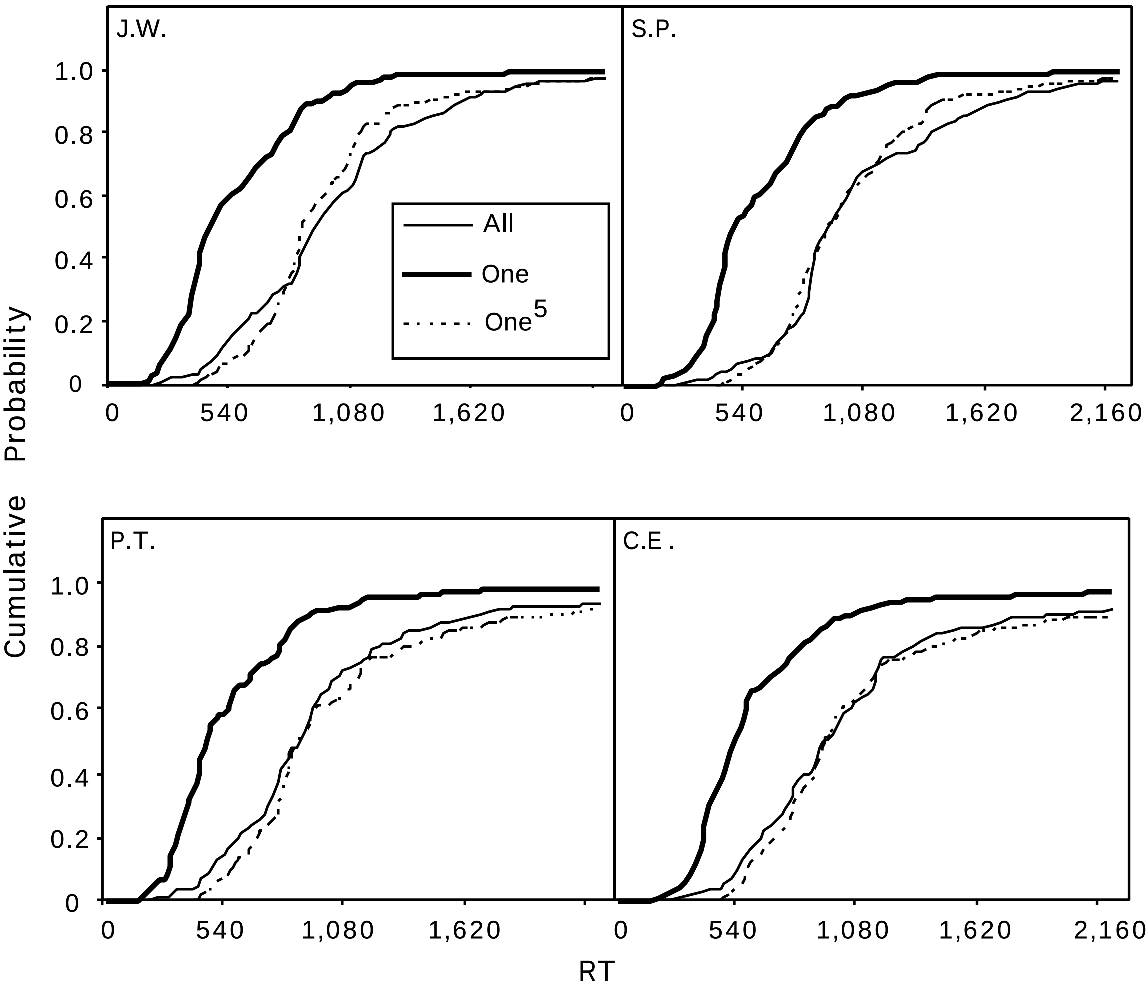


Figure 12. The cumulative response time distributions for the four participants taken from the last session. The One and All distributions were empirically obtained. The One$^5$ distribution is obtained by raising each point along the One distribution to the fifth power, and represents the analytic model's predicted cumulative distribution for the All task. Violations of this analytic model occur when the All task distribution is shifted to the left of the analytic model's distribution. Such violations occur for the fastest half of response times for all four participants.

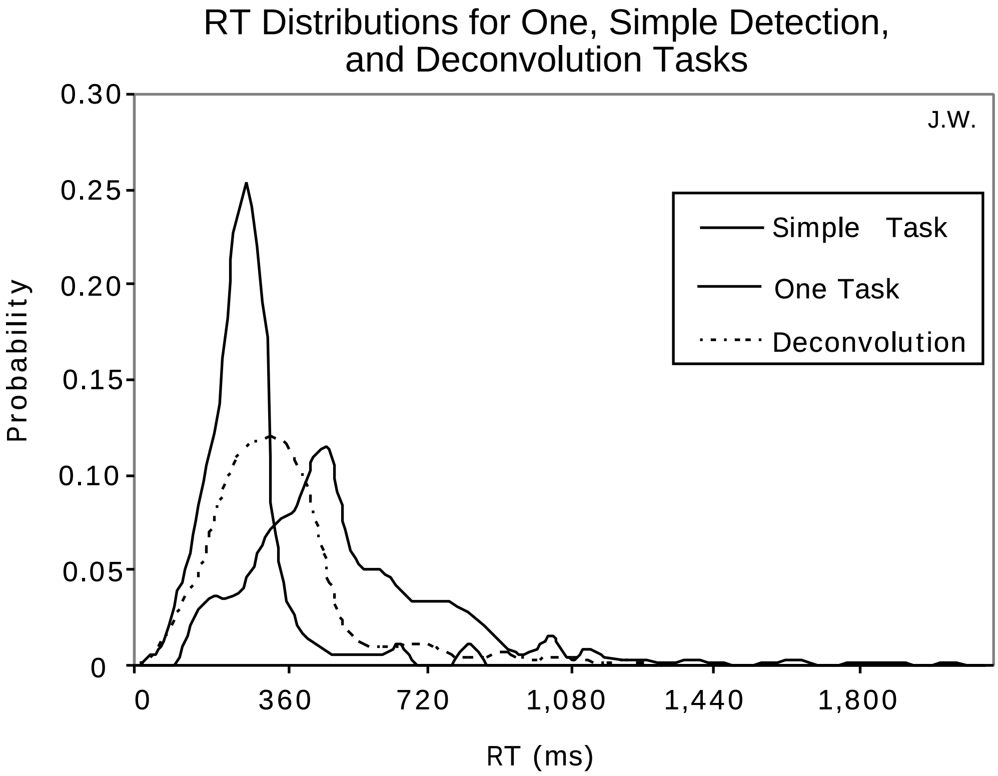


Figure 13. A demonstration of the Fourier transformation method of response time decomposition for C.H.'s data. By deconvolving the simple detection task distribution from the One task distribution, the distribution labeled "deconvolution" is obtained. This represents the distribution of times required for the comparison process of the One task, removing processes associated with the simple detection task. This decomposition of the One task allows us to sample five times selectively from the comparison process, and combine (convolve) these times with those associated with simple input and output processes.

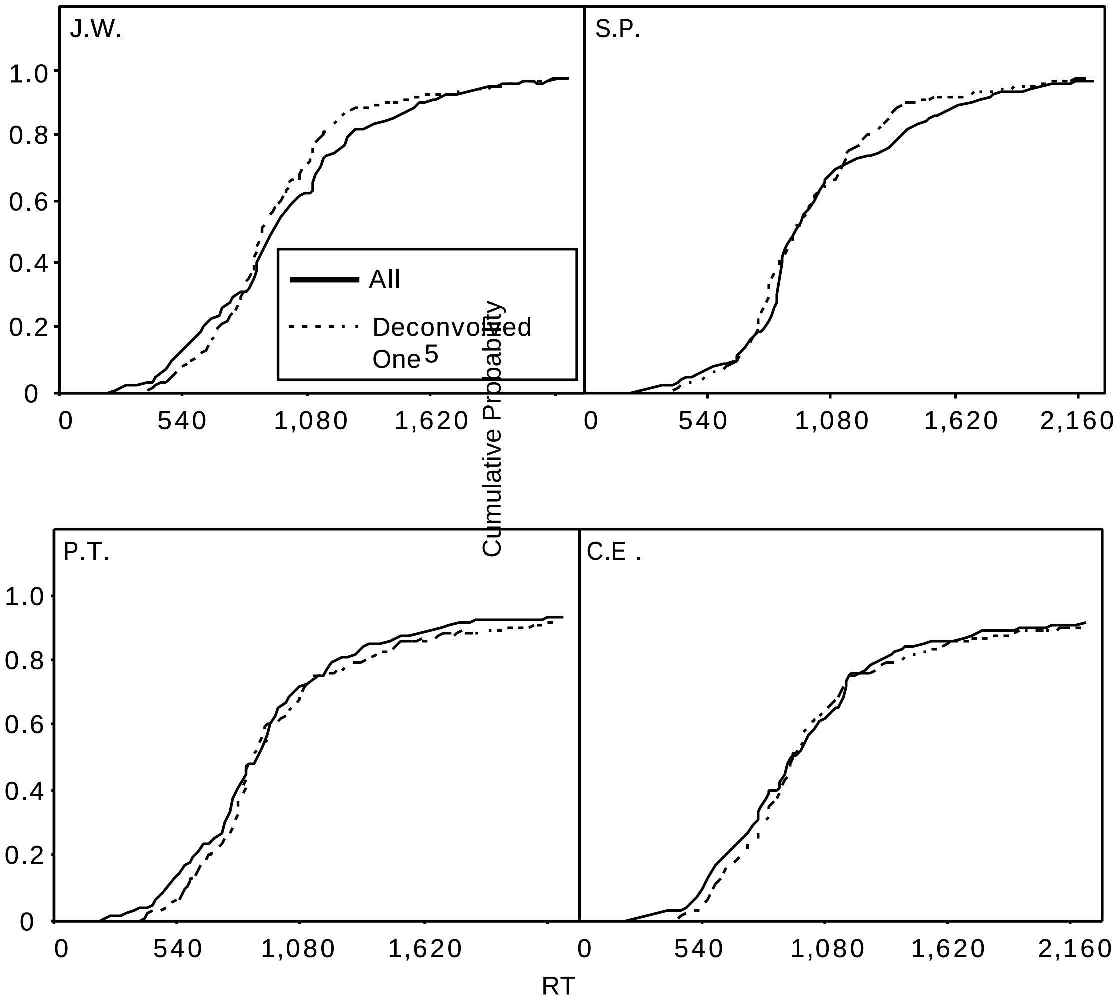


Figure 14. The newly derived analytic model's predictions are superimposed with the participants' performances on the last experimental sessions. Despite relaxing the assumption of independent sampling, violations of the analytic model are still found for three of the four participants.