Journal of Experimental Psychology:
Human Perception and Performance
1993, Vol. 19, No. 2, 331-351

Copyright 1993 by the American Psychological Association, Inc.
0096-1523/93/$3.00

# What Enumeration Studies Can Show Us About Spatial Attention: Evidence for Limited Capacity Preattentive Processing

Lana M. Trick and Zenon W. Pylyshyn

Subitizing, the enumeration of 1–4 items, is rapid (40–120 ms/item) and accurate. Counting, the enumeration of 5 items or more, is slow (250–350 ms/item) and error-prone. Why are small numbers of items enumerated differently from largenumbers of items? It is suggested that subitizing relies on a preattentive mechanism. Ss could subitize heterogeneously sized multicontour items but not concentric multicontour items, which require attentional processing because preattentive gestalt processes misgroup contours from different items to form units. Similarly, Ss could subitize target items among distractors but only if the targets and distractors differed by a feature, a property derived through preattentive analysis. Thus, subitizing must rely on a mechanism that can handle a few items at once, which operates before attention but after preattentive operations of feature detection and grouping.

This article has two goals. The first is to answer a longstanding question about enumeration: Why are small numbers of items enumerated differently from large numbers of items? The second is to show that studies of enumeration can reveal something new about vision, namely that there is a limited-capacity stage between the unlimited-capacity preattentive and one-at-a-time attentive stages of visual analysis.

Suppose that an observer had to enumerate the dots in the displays in Figure 1. He or she might notice that there is something qualitatively different between the experiences of enumerating the dots in Figures 1a and 1b. In Figure 1a, enumeration seems effortless and immediate; the observer simply "sees" how many dots there are, and moreover, is certain of his or her response. In Figure 1b, the observer might notice that enumeration seems slow and laborious. If this observer is like most adults, he or she may be conscious of grouping the dots into clusters of 2–4, moving from cluster to cluster, finding the number of dots in each cluster, and then adding this number into a running total (cf. Van Oeffelen & Vos, 1982; Warren, 1897). At the end of this laborious process, the observer might still feel less confident of the estimate than that for Figure 1a, even though more time was spent on Figure 1b. In fact, errors become more common when there are large numbers of items, so any lack of confidence is justified (cf. Jensen, E. Reese, & Reese, 1950; Kaufman, Lord, T. Reese, & Volkmann, 1949).

When enumeration latency is measured as a function of the number of dots in the display, a discontinuity in slope is evident. When there are small numbers of items (as in Figure 1a), the slope of the latency function is shallow; for adults, each additional item adds between 40–120 ms in the 1–4–item range (e.g., Akin & Chase, 1978; Klahr, 1973a; Oyama, Kikuchi, & Ichihara, 1981). When there are five or more items in the display, (as in Figure 1b), the slope is large, between 250–350 ms. This discontinuity occurs even when the threshold exposure required for a certain level of accuracy is measured instead of reaction time. In threshold experiments, dots are shown for 5–10 ms only, and the dependent variable is the stimulus onset asynchrony between dots and a pattern mask needed to ensure that subjects are 50% accurate in their report of the number of items. The absolute slopes are lower in threshold experiments. For example, in a typical threshold experiment the slope in the 1–4–item range is 4–10 ms/item (Oyama et al., 1981) and 1.9 ms/item if a subjects only have to discriminate $n$ from $n + 1$ (Sagi & Julesz, 1984). Similarly, the slope for five or more items drops to 60 ms/item (Oyama et al., 1981). Nonetheless, the discontinuity in slope between small and large numbers of items remains.

Scores of studies (dating from that of Jevons, 1871) document differences in the ease, accuracy, and rate of enumeration between small and large numbers of items. These results are usually interpreted as evidence that there are two enumeration processes. One process is specialized for small numbers of items and is fast, accurate, and effortless. This process was called *subitizing* by Kaufman et al. (1949).[1]

Lana M. Trick and Zenon Pylyshyn, Department of Psychology, University of Western Ontario, London, Ontario, Canada.

We thank Jim Enns, Ron Rensink, and Debbie Aks for helpful comments on the early drafts of this article. We also thank Robert Frick, Nancy Kanwisher, and two anonymous reviewers for their suggestions.

Correspondence concerning this article should be addressed to Lana M. Trick, who is now at the Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4.

[1] Kaufman, Lord, Reese, and Volkmann (1949) originally used *subitizing* to refer to the rapid, accurate enumeration of items in a short-duration display. Jensen, E. Reese, and Reese (1950) amended this definition to encompass both short-duration displays and displays that remained on until the subject responded. They noted that there are problems associated with using limited exposure duration as part of the operational definition of subitizing. In particular, they claim that if subitizing is by definition rapid, accurate enumeration with short-duration displays, and counting is by definition a slow, accurate process that occurs with long-duration displays (see Warren, 1897), it is impossible to look at subitizing and counting using the same presentation conditions. It would be possible to compare subitizing and estimation (an inaccurate form of enumeration used for high numbers of items) with limited exposure durations only. If the 200-ms exposure duration

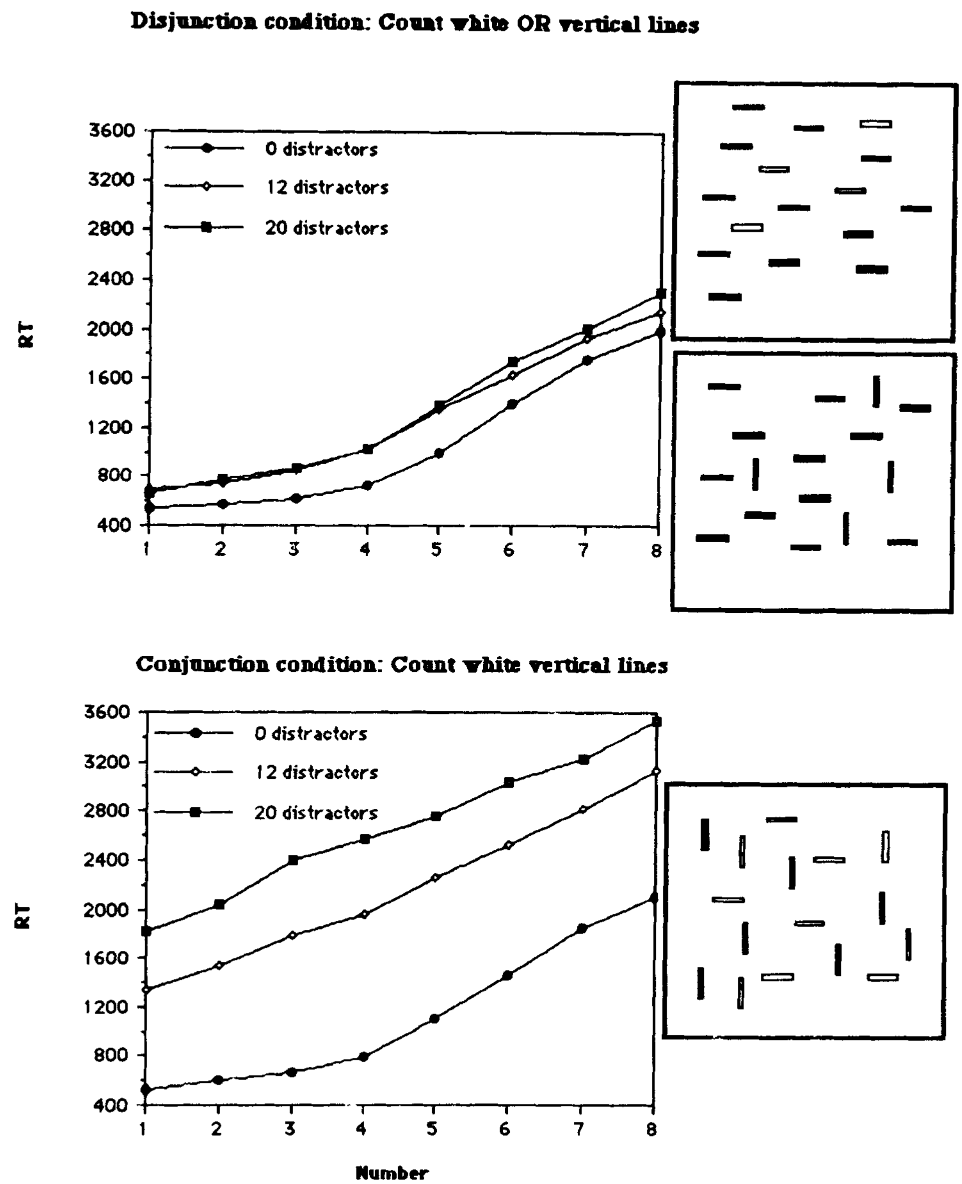


*Figure 1.* Enumeration latency as a function of number (idealized). (RT denotes reaction time.)

The other process can handle large numbers of items but is slow, effortful, and error-prone. This process has been called *counting*, which is awkward, because counting is the colloquial term for the enumeration task. In this article, we will reserve *counting* for the process and *enumeration* for the task.

The discontinuity in slopes for enumerating small and large numbers of items produces an elbow in the reaction time function. This elbow is taken to be the boundary between the subitizing and counting range. Estimates of the subitizing range vary between 1–3 and 1–7 items from study to study, depending on the particular paradigm, accuracy criteria, display contrast, and statistical technique used to calculate the subitizing range. There are individual differences among subjects in how many items can be subitized even within the same study, however (e.g., Akin & Chase, 1978; Jensen et al., 1950), and as a result, the transition in slopes is seldom as sharp as Figure 1 would indicate. Averaging data across subjects, some who subitize to 3 items, some who subitize to 4 items, some who subitize to 5 items, and some who subitize to 6 items, produces a rounding in the upward turn of the function. For convenience, we refer to the subitizing range as *1–4*, as have many others (cf. Aoki, 1977; Atkinson, Campbell, & Francis, 1976; Oyama et al., 1981; Simons & Langheinrich, 1982). We are not interested in making strong arguments that the subitizing range is exactly 4, however. That sort of debate is pointless given that there is evidence for individual differences among subjects.

recommended by Kaufman is used for five or more items, the error rate is very high, for example, 70% or more after six items (Mandler & Shebo, 1982). When we count in day-to-day situations, our accuracy is considerably better.

Although enumeration has been studied for more than 100 years, no one has satisfactorily solved the mystery of why small numbers of items are enumerated differently from large numbers of items. Why is there a slope discontinuity in the enumeration function? Simple retrieval of number names takes no appreciable extra time after 4 items (e.g., Mandler & Shebo, 1982). Nor do addition (+1) latencies suddenly jump after 4 items (Parkman & Groen, 1972). We have a very fast, accurate process for enumeration—the subitizing process. Why can we not subitize any number of items?

The modern study of enumeration began when Hunter and Sigler (1940) and Saltzman and Garner (1948) observed that the visual characteristics of the display affect enumeration. Since then, two theories have emerged. Both theories are inadequate for explaining the flexibility of human enumeration, however. For example, according to density theory (Atkinson et al., 1976), the reason we can subitize only small numbers of items is that there are neural units that selectively respond to low-frequency gratings of a particular phase. These units could be used for enumeration of small numbers of items (sparse displays with few contours), though no such units exist for higher spatial frequencies. The problem with density theory is that humans enumerate objects, not contours. Objects have a variable number of contours, and consequently contour density is a poor heuristic for number even in the 1–3 range. Yet humans are capable of subitizing complex objects, even three-dimensional block figures that partly occlude one another (e.g., Akin & Chase, 1978).

Similarly, according to pattern theory (Mandler & Shebo, 1982), humans can subitize only small numbers of items because subitizing involves canonical pattern recognition: 1 item forms a point, 2 items can be connected by a line, and 3 items typically fall into a triangular configurations. Beyond 3 items, pattern cannot be used as enumeration heuristic because there is no characteristic pattern that each number of items falls into. The problem with pattern theory is that subitizing occurs even when the items do not fall into the prescribed patterns. For example, in Frick's (1987) study, subjects enumerated up to 20 items presented in a row. Although items always fell in the linear configuration, the configuration for 2 items according to pattern theory, subjects enumerated with accuracy. Moreover, the slope discontinuity between small and large numbers of items, a trademark of the change from subitizing to counting, was evident. In fact, subjects enumerate items in atypical patterns of 3 as quickly and accurately as they enumerate triangular patterns (Trick, 1987).

Density and pattern theory fail because of a lack of generality. Humans can enumerate not only single contours, but objects made up of multiple contours or even illusory contours; humans can enumerate not only items that fall into typical patterns, but ones that fall into atypical configurations as well. In fact, humans can enumerate objects that move and change their retinal properties from moment to moment and even enumerate objects in a background of distractor items. Humans can enumerate almost anything they see, and any theory of enumeration must be comprehensive enough to take this into account. The enumeration of visual items that occupy different locations in space needs to be considered as a visual–spatial task and needs to be understood in the context of visual–spatial processing in general.

In the following sections, we discuss subitizing and counting in the context of a general theory of vision inspired by Marr (1982), Treisman and Gelade (1980), Pylyshyn (1989), and Ullman (1984), and we then present a series of experiments that suggest that subitizing can be explained only by virtue of a limited-capacity mechanism that operates after the spatially parallel processes of feature detection and grouping but before the serial processes of spatial attention.

## Enumeration Within a General Theory of Vision

Visual processing has been thought to involve two stages. The first is a spatially parallel preattentive stage, where analyses occur at every location at the same time. Feature registration and grouping occur at the preattentive stage. Features—properties such as brightness, color, orientation, and curvature—are computed at every location in the image. Discontinuities are registered, and, according to Marr (1982), assigned a place token. For example, a green area next to a black area would be assigned a place token. Then, feature clusters are formed by grouping the place tokens associated with the discontinuities on the basis of gestalt properties of proximity, similarity, good continuation, and common fate. For example, on a black background, several green points might be grouped to form a line, and the line of points would be assigned a place token. Because feature registration is spatially parallel, items that differ from other items on the basis of a feature can be detected in a time independent of the number of items in the display; for example, a red item pops out among green (Treisman & Gelade, 1980). Because of the tendency to group adjacent similar discontinuities, effortless texture segregation occurs (Beck, 1982; Julesz, 1984). Subjects see a group of red items in a group of green as being surrounded by an implicit contour that can define a shape, without having to scrutinize the display area by area.

The second, attentive stage of visual analysis is spatially serial. Analyses are performed one item and one location at a time by moving a processing focus through the image. This focus can be understood as a spotlight (e.g., Posner, Snyder, & Davidson, 1980), or the locus of the majority of the attentional resources in a gradient (LaBerge & Brown, 1987). Regardless, this focus performs detailed perceptual analysis. In particular, the attentional focus is needed for integrating parts of an object into a whole (Treisman, 1988; Ullman, 1984), computing spatial relations such as *inside* and *connected* (Ullman, 1984), and integrating different features of an object into an object description (an episodic representation Kahneman and Treisman, 1984, called the *object file*). After the object file is complete, this description is matched to memory representation so recognition and naming can occur.

This way of understanding vision assumes that analysis occurs either all at once or one item at a time. As a result, there is no way to explain subitizing, where small numbers of items are treated differently from large numbers of items. What causes the slope to increase from 40–100 ms/item to 250–350 ms/item? We argue that the discontinuity occurs because a preattentive mechanism can register the presence of each of a small number of items, as long as the number of items in the display does not exceed the capacity of the mechanism. If there are more items than can be handled by this mechanism, enumeration requires several additional operations. In particular, however, enumeration of larger numbers of items requires moving the attentional focus from area to area in the display.

Thus, we argue that subitizing reflects the operation of an intermediate stage between preattentive and attentive analysis, as shown in Figure 2. This intermediate stage performs individuation. Individuation is necessary to distinguish between different tokens of same type, for example, a given black dot from all other black dots. Enumeration by its nature requires individuation. For example, in order to enumerate three black dots, it is necessary to be aware that there is one black dot, which is a separate entity from another black dot, which in turn is a separate entity from yet another black dot. Each item has to be considered separate and distinct for the final number estimate to be correct. It is imperative to keep each item being enumerated distinct from those already enumerated and those yet to be enumerated.

More generally, however, individuation is necessary every time the position of the attentional focus is changed from visual item to visual item, an operation Ullman (1984) called *indexing*. Indexing is like driving. To deliberately "drive" attentional focus from one particular item to another, one needs to know both where one is and where one is going. At the very least, this requires a way of referring to the destination. For example, suppose that the attentional focus is currently at Point a in Figure 3. The task is to move the attentional focus point b. How could this be accomplished?

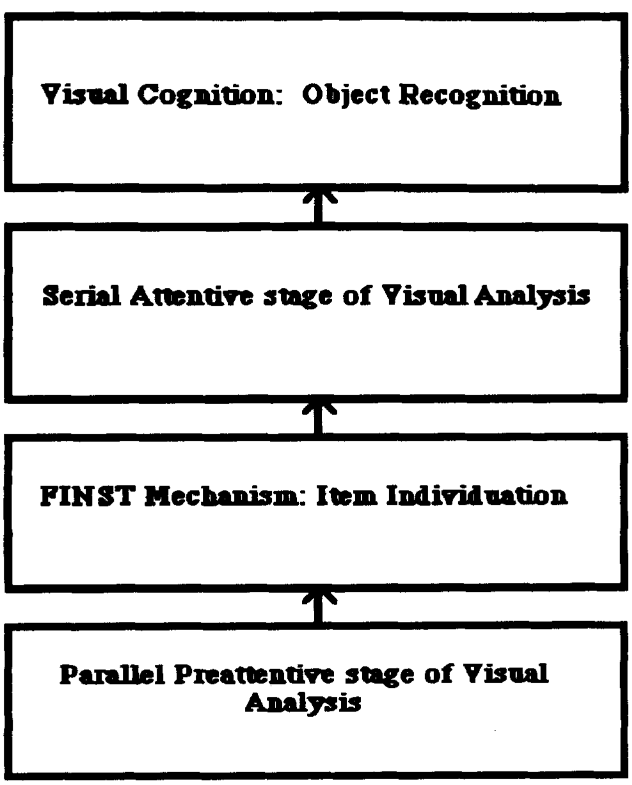


*Figure 2.* The FINST (fingers of instantiation) mechanism in visual processing.

One possibility is for the processor to be sent to a location defined by the properties of the item, for example, INDEX (small black dot). This strategy would not work, however, if there were more than one token of the same type in the image, that is, if there were more than one small black dot. In this example, the processor might be as likely to land at Point c or d if a property-based method of address was used. Moreover, in the real world, the retinal properties of an item might change from moment to moment as a result of changes in lighting or projection or changes in the item itself. A dot might on its own change color, brightness, or size, for example. For these reasons, the strategy of using properties as arguments in the index operation is doomed. There needs to be a way to refer to an item, maintaining its identity, even if the item changes its properties: It is the same one, although it used to be little and black, and now it is big and gray.

Another possibility is that a coordinate grid be overlaid and the index operation take as its argument a retinal coordinate such as INDEX (25, 35). If Treisman and Schmidt (1982; cf. Mozer, 1989) are correct, however, there is little detailed position information available at the preattentive level. That is why miscombinations of features, illusory conjunctions, occur unless attention is focused on the location of the target item. Subjects might falsely believe that they saw a red circle and a green square when in fact they saw a green circle and a red square, for example. Furthermore, the retinal coordinates of the item change with eye, head, or body movement. For example, Item b might fall at the same retinal location as Item e if the person was to move their eyes to the left. Finally, objects move by themselves. Thus, the position of Item b might change even if eye position was constant. To index on the basis of retinal coordinates is to risk sending the attentional focus to a retinal location that either no longer houses an item or houses an item different from the one intended. There needs to be a way to refer to an item independently of its location so that its identity will be maintained even though its position changes: It is the same one, though it used to be at (25, 35) and now it is at (63, 90). Although at any given moment each item will be in different retinal position fromn the other visible items, it would be a bad idea to refer to items by their position because item positions change.

If properties and locations cannot be used to refer to items for indexing purposes, what can? There needs to be some way of naming the individual items, assigning mental reference tokens, in the same way we assigned physical reference tokens by naming the points a, b, and c. It is not enough to simply individuate one item, however. If Ullman (1984) is correct, and spatial relations such as inside, connected, and ahead of are computed using the attentional focus, sometimes several items must be individuated at once. It is necessary to keep several selected target items distinct not only from distractor items, but from each other.

For example, in order to determine whether the points we have labeled Y and Z in Figure 3 are connected by Contour X, it is necessary to consider Y and Z as separate entities, not to mention the contour between them. If the attentional focus can only be in one location at a time, as many seem to believe (e.g., Posner et al., 1980), there must be a way to preattentively individuate several items so that the attentional focus can be used to compute spatial relations.

Pylyshyn (1989) proposes that we have a small number of mental reference tokens, *FINSTs* (fingers of instantiation), in order to perform this task. FINSTs, like human fingers, allow one to specify "that one" without explicitly stating item properties or locations. Moreover, FINSTs, like pointing fingers, give one information about where an item is. To reiterate, two things are important. First, there must be a way to name items, assign reference tokens, in order to refer to the items. Second, it is important that these tokens carry information about where the associated items are in the image, in the same way that pointer variables carry information about the memory locations of their associated variables in computer languages such as Pascal or C. Without the ability to name feature clusters, so they can be referred to independently of their properties or locations, there would be no way to ensure that the focus of attention would arrive at the intended destination. Without some way of knowing where selected items are from moment to moment in the retinal image, there would be no way to send the attentional focus to the correct retinal location. Deliberate strategic control of the attentional focus would be impossible.

According to Pylyshyn (1989), there are only a small number of FINSTs. This limitation seems reasonable because FINSTs are used in selecting items for attentional processing; it makes little sense to select every item in the image. At the same time, in order to compute spatial relations and perform tasks such as multiple target tracking, in which subjects are required to keep track of four or five items in a field of as many identical distractors, when targets and distractors are all set into rapid independent movement (Pylyshyn & Storm, 1988), it is necessary to select more than one item at a time. Nonetheless, it may be advantageous not to individuate more than a very small number at once. Simultaneously tracking a large number of items involves solving a difficult problem—the motion correspondence problem, determining which item in one frame of motion corresponds to which item in the next (e.g., see Ullman, 1981). As the number of items to be tracked increases, there is a combinatorial explosion in the number of possible frame-to-frame correspondences between items. If only a few items are selected for tracking, the correspondence problem could be simplified.

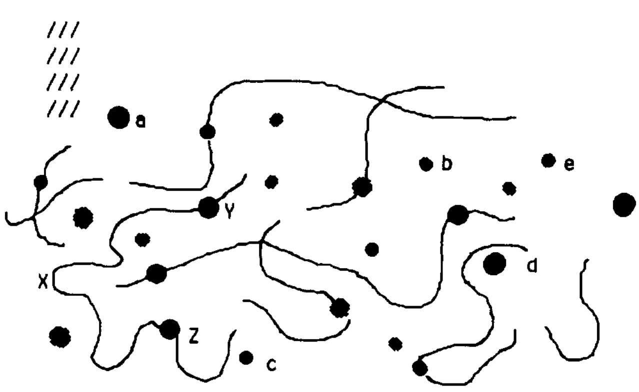


*Figure 3.* Moving the attentional focus between items, computing spatial relations, and grouping at different levels of resolution.

FINSTs may be automatically assigned to every feature cluster in a very austere image, an image with few discontinuities. Similarly, luminance transients may automatically attract FINSTs (though color transients do not; see Pylyshyn & Burkell, 1990). In most cases, however, FINSTs have to be assigned selectively in response to goals. Only a few feature clusters in a typical image will be FINSTed. There are three ways in which this selective assignment might take place. First, subjects can choose what resolution to work at. For example, in Figure 3 FINSTs may be assigned to place tokens corresponding to the diagonal lines, the horizontal lines made up of the diagonals, or vertical lines made up of horizontals. Consequently, it is possible to enumerate at any resolution, though it may be easier at some resolutions than others.

Second, subjects may be able to choose to assign FINSTs only to items with certain features. Thus, subjects may choose to FINST all the red items and ignore the green. In fact, this prediction was borne out in guided search studies that show that subjects sometimes look for conjunctions of features by checking only the items with one of the relevant features. For example, a person might search for a red O among red Ns and green Os, by checking only red items (e.g., Egeth, Virzi, & Garbart, 1984; cf. Wolfe, Cave, & Franzel, 1989). This sort of selection should only be possible if targets and distractors differ by a property that mediates pop out in search, however—a property derived during preattentive analysis. The reason that FINSTs can be assigned selectively in this case would be that the information needed to determine which items to FINST is available before the stage of individuation. If spatial attention were required first to determine which items were targets, FINSTs could not be assigned selectively to targets. The information needed to determine which items to select would not be available until after it was needed.

Third, FINSTs can be released in response to a goal after processing is complete so that they can be reassigned to other items. This is necessary to explain the processing of displays with large numbers of items, as in enumeration of 20-item displays, for example. This is also necessary to explain how a number of different spatial relations can be derived from a complex display. For example, after finding out if one's horse is ahead of the others in the horse race, one might also want to find out if one's horse is on the inside lane in relation to the track.

Why can humans subitize only small numbers of items? The reason is that the system that individuates feature clusters by binding them to mental reference tokens has limited resources; there are only a small number of reference tokens (FINSTs). Consequently, the system is spatially parallel but nonetheless limited capacity. If there are more items than reference tokens, a different process must be used, one that involves assigning and reassigning FINSTs and moving the focus of attention from area to area in the display, adding the number of items in each area into a running total. This latter process can be thought of as grouping and adding, the process that most adults seem to use when they enumerate large numbers of items (cf. Klahr, 1973a; Van Oeffelen & Vos, 1982).

Only theories that talk about mechanisms capable of handling small groups of items at once can explain the slope discrepancy in the enumeration function between small and large numbers of items. FINST theory is not the only one that talks about the ability to deal with small numbers of items at once, however. Ward and McClelland (1989) suggest that it is possible to search for several identical conjunction targets simultaneously, but they give little consideration to the problem of individuation.

Moreover, FINST theory is not unique in its concern with individuation and enumeration. Mozer (1989) argues that preattentive analysis is capable of processing every item to the point of identification, but problems occur because the identity information comes without detailed spatial information. He predicts that item homogeneity should interfere with enumeration because items with the same properties (tokens of the same type) cannot be individuated by their spatial locations because little spatial information is available early on. His work does not predict where in the number range homogeneity should have an effect. There is nothing to suggest homogeneity would not interfere even with the enumeration of two items (well within the subitizing range). Although Mozer discusses enumeration, he considers subitizing irrelevant to the issue and tries to prevent it (e.g., Experiment 1). Mozer's account cannot explain why homogeneity has no effect on latencies for fewer than seven items (e.g., Frick, 1987). Moreover, his strategy of using either properties (type) or spatial locations to individuate items would fail in real-world situations where items move and change their retinal properties from moment to moment.

Thus, FINST theory differs from others in suggesting that there are preattentive limitations in the number of items that can be individuated that arise because there are a limited number of mental reference tokens, or FINSTs. FINST theory is unique in positing a relationship between this capacity limitation and multiple target tracking on one hand (Pylyshyn & Storm, 1988) and subitizing on the other.

## Overview of Studies

Our goal is to test one prediction of the FINST hypothesis, namely that there is a limited-capacity stage, the basis for the subitizing process, that precedes the operation of spatial attention. In particular, we will be trying to show that complex displays can be subitized, as long as the information needed to resolve the items to enumerate is available before attentional analysis. The item individuation mechanism must operate before spatial attention if it is to provide reference tokens for the attentional focus, as we suggest. Conversely, subitizing should not be possible if spatial attention is required to resolve the items to be enumerated.

Other theories either predict that subitizing should never occur, regardless of the enumeration task, or subitizing should always occur, regardless of the enumeration task. For example, Ullman (1984) claims a combination of elementary attentional operations are combined to make up a program, a "visual routine" to compute number. Yet there is nothing in his model to predict subitizing.

In contrast, feature integration theory (e.g., Kahneman & Treisman, 1984; Treisman & Gelade, 1980) could be interpreted to predict that subitizing should always occur, regardless of the attentional demands of the enumeration task. This prediction is based on two tenets from classic feature integration theory. First, features come with little or no location information (Treisman & Schmidt, 1982). As a result, there would be no way to individuate different items that share a particular feature if the items were presented with distractors. Notice that two things are important here. Targets must be kept distinct from each other, and targets must be kept distinct from the distractors. Second, object files can be defined only one item at a time, using the attentional focus (Kahneman & Treisman, 1984). Thus, although items could be individuated by object files, these files must be created one at a time. There is nothing in feature integration theory that could explain why 1–4 items would receive different processing from 5 or more items; subitizing and counting could be explained only by processes that operate after both preattentive analysis and the attentive processes that define object files. Consequently, there is no reason to suspect that attentional requirements of the task would have an effect on whether subitizing would occur from feature integration theory. Feature integration theory would predict slope discontinuities in the enumeration function whether feature integration is necessary or not, though the theory would predict that overall enumeration latencies would be higher if attentional processing were required.

Only a theory that appeals to a limited-capacity preattentive mechanism would predict that the attentional requirements of the enumeration task would affect whether subitizing occurs. We predict that differential enumeration of small and large numbers of items should only occur if spatial attention is not needed to resolve items as wholes, or distinguish targets from distractors.

In each experimental trial subjects were shown a display, required to say how many items there were, and then type in their response. The dependent measure was vocal reaction time, the amount of time required to correctly say how many items there were. We will be looking for situations in which subitizing does not occur. How will we know when this happens? The trademark of the change from subitizing to counting is the increase in slope in the latency function after 3 or 4, which trend analysis registers as a deviation from linearity. If an experimental manipulation gets rid of this deviation, there is evidence that the same process is being used for both small and large numbers of items. Because trend analysis cannot distinguish deviations from linearity that result from an increase in slope from those caused by decreases or plateaus, it was necessary to check the direction of the trend at the point in the number range where it first became significant (1–3, 1–4, 1–5, etc). In addition, because averaging across subjects could obscure discontinuities in slope as a result of different subjects having different subitizing ranges, trend analyses were performed on each subject's data independently.[2] Analyses were also performed on the averaged latencies, in which each case in the analysis represented 1 subject's average enumeration latency for a certain number and condition.

Finally, as an additional check, regressions were performed to calculate slopes. In all experiments, the majority of the subjects subitized to 4; the deviation from linearity did not emerge until the 1–5 range was analyzed. The subitizing range for individual subjects varied between 1–2 and 1–6, however. Because all but 1 subject subitized at least to 3, the subitizing range was considered 1–3 for purposes of regression. This was done to avoid unduly inflating the subitizing slope with latencies in which subitizing did not occur. Similarly, the counting range was considered 5–7 to avoid contaminating the counting slope with end effects. End effects occur in reaction time studies when subjects have knowledge of $n$, the maximal number of items that they will be shown in a display, and begin to enumerate $n$ items almost as fast or faster than they enumerate $n - 1$, perhaps because they adopt a strategy of guessing $n$ whenever there are a large number of items in the display (cf. Atkinson et al., 1976; Mandler & Shebo, 1982). In addition, slopes were calculated from the data of the majority of subjects when there was evidence that some subjects were performing the task differently from others (i.e., subitizing when most counted, or vice versa). Including atypical subjects in the analysis would simply serve to decrease the fit of the function and inflate or deflate the slopes. Moreover, it is probably futile to try to interpret slopes averaged across subjects who are using different strategies (cf. Siegler, 1987).

## Resolving Items as Wholes

### *Experiment 1*

Usually when humans enumerate they count objects rather than points of light. Objects may be defined by many contours or intensity discontinuities, and at times objects may partially occlude one another. Observers seem to be able to subitize anyway (e.g., Akin & Chase, 1978). How is this possible? According to Ullman (1984), many of the properties that define multiple contour objects as wholes require attentional processing. If subitizing can only make use of preattentive information, how is it possible to subitize multiple contour objects? We argue that it is not always possible to subitize multiple contour objects, and in cases where preattentive processing is misled into grouping contours from different objects, or assigning the same place token to several different objects, subitizing becomes difficult.

When is preattentive information an unreliable cue to the number of objects? According to Marr (1982), place tokens are assigned to intensity discontinuities, and then tokens are grouped by proximity, similarity, and common fate. Thus, discontinuities that are close together and similar in contrast, orientation, depth, brightness, size, or motion may be grouped together and assigned a place token. Then a group of place tokens may itself be assigned a place token, and the process could repeat. Thus, for purposes of preattentive vision, each place token corresponds to a unit at a certain level of analysis.

In most cases nearby and similar contours belong to the same item, and consequently the number of "units" derived by preattentive vision would correspond to the number of items in the display. For concentric displays, however, the closest and most similar contours come from different items. If all the contours radiate around a common focus, there might be a tendency to group all of the contours into one unit. Consequently, the number of units derived by preattentive grouping processes would not correspond to the number of concentric items.

Given this expectation, it is interesting that one of the few studies that failed to show clear evidence of subitizing had subjects enumerating concentric circles (Saltzman & Garner, 1948). There is little evidence of slope discontinuities, trademark of the change from subitizing to counting, from their concentric circle data in Figure 4. In contrast, discontinuities are apparent in their dot enumeration study, presented in Figure 5.

Saltzman and Garner's results are suspect for several reasons, however. First, trend analyses were not performed, and means and standard deviations for the latencies were not listed, so it is hard to say whether the apparent slope discontinuities are significant. Second, Saltzman and Garner gave their subjects ample opportunity to use cues other than number to decide the cardinality of the display in their concentric circle study. They set up the displays so that the largest ring was always 14° in diameter, and the distance between rings was the radius of the smallest ring in the display. This would make it possible for subjects to avoid the enumeration task, and with practice, simply make the number judgment on the basis of the distance between rings, or the diameter of the innermost circle. Furthermore, they gave their subjects many exposures to the same stimuli, and with repeated exposure subjects may begin to use their memory for form to avoid enumerating the items (Mandler & Shebo, 1982).

The present study had two objectives. The first was to replicate Saltzman and Garner's concentric item result, performing the appropriate statistical analyses and setting up displays in such a way that the cardinality could not be decided except by counting or subitizing the items. Moreover, we wanted to ensure that grouping contours by relative proximity was possible, given that grouping by proximity

[2] In the more recent experiments (Experiments 2, 3, and 5), efforts were made to ensure that each cell of the analysis had the same number of cases because unequal cell variances were typical; a given subject might have a standard deviation of 50 ms when counting to 1 and 300 ms when counting to 7. When unequal cell variances are accompanied by unequal cell sizes, $F$ statistics, used in trend analysis, are exaggerated. Thus the probability of Type I error increases (Milligan, Wong, & Thompson, 1987). In order to keep the number of trials per cell equal, error trials were readministered until subjects got them right. (The appearance of re-administered trials was disguised, either by making the display a mirror image rotated 90°, changing the side of the display that the items appeared on, or changing the starting point for counting.) To prevent extreme latencies from unduly affecting the means, one outlier per cell was dropped.

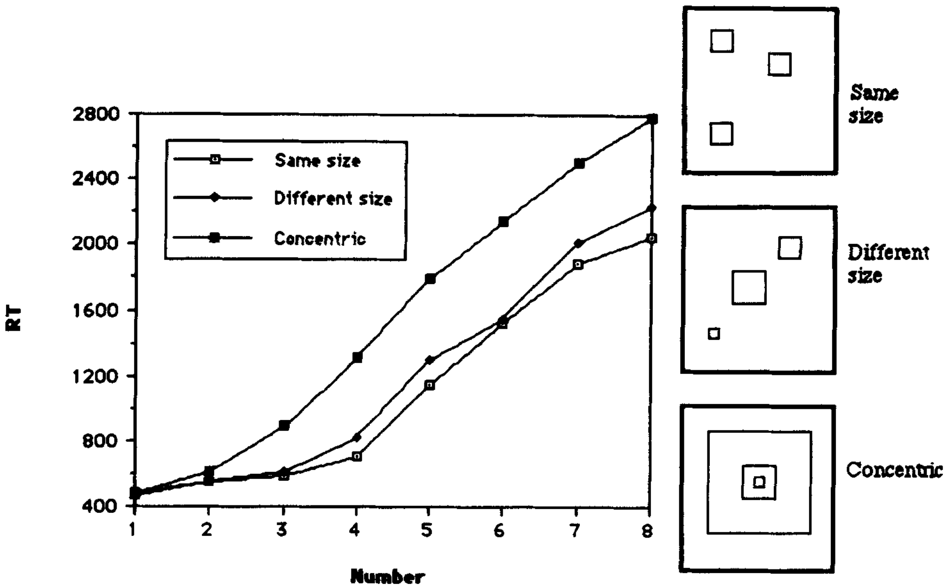


*Figure 4.* Average latencies to enumerate concentric rectangles as opposed to rectangles of homogeneous or heterogeneous size. (Standard errors for the means varied between 10.4 and 182.7, with larger standard errors associated with larger enumeration latencies. RT denotes reaction time.)

seems to be a natural component in the enumeration process (Van Oeffelen & Vos, 1982), although subjects appear to be capable of subitizing and counting equispaced items (Atkinson et al., 1976; Frick, 1987). Finally, we wanted to ensure that the result would replicate even if the items were presented within the foveal area of the retina, to ensure that Saltzman and Garner's results were not an artifact of their very large (14°) display size.

The second objective was to discover why subjects seem capable of subitizing dots but not concentric circles. Saltzman and Garner's dot enumeration task differed from their concentric circle task in three ways. First, circles seem more complex than dots; circles are larger and are defined by a contour surrounding an interior that is the same color as the exterior of the item. Dots, on the other hand, are so small that they might conceivably be registered as edge points; they are also uniformly colored and different from the background. Second, all the circles were of different sizes in the concentric task, whereas all the dots were the same size in the dot enumeration task. Finally, the concentric circles were one inside another and shared a common focus, whereas the dots were dispersed through the display. We will be trying to establish which of these factors caused Saltzman and Garner's results.

Consequently, in this experiment there were three conditions. In the same-size condition, subjects were required to enumerate rectangles of the same size distributed across the screen. The different-size condition was similar except at

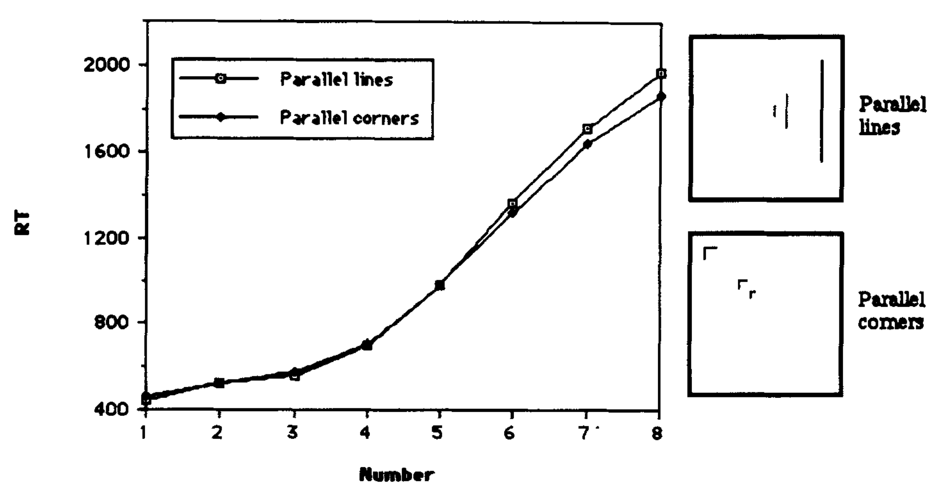


*Figure 5.* Average latencies to enumerate parallel lines and parallel corners. (Standard errors for the means varied between 21.6 and 121.2, with larger standard errors associated with larger enumeration latencies. RT denotes reaction time.)

least one of the rectangles was a different size from the others. Finally, in the concentric condition, subjects were required to enumerate concentric rectangles. If Saltzman and Garner's concentric results came about simply because subitizing only occurs for dots rather than multiple contour objects, subitizing should not be evident in any of the three conditions. If the result came about because concentric items are necessarily of different sizes, there should only be evidence of subitizing in the same-size condition. Finally, subitizing should be evident in all but the concentric condition if subitizing did not occur in Saltzman and Garner's study because the items were concentric.

We expected that subitizing would be evident in same-size and different-size conditions but not the concentric condition. Why would this be the case? Consider a display of white outline rectangles on a black background. Preattentive processing would deliver a representation in which illumination discontinuities were grouped, primarily by proximity in this case. Each cluster would be assigned a place token, and this place token could in turn be FINSTed. When objects are distributed throughout the display, as in the same- and different-size conditions, these clusters would correctly reflect the number of objects. Contours that are the closest together typically come from the same object. Thus a FINST could be assigned to each cluster, and subitizing could carry on as usual.

If the rectangles are concentric, this procedure would not be possible. The contours that are closest together are inevitably from different objects when items are concentric, and moreover, these immediately adjacent edges and corners also have the same orientation. Thus, there would be a tendency to group the wrong clusters on the basis of both proximity and similarity. Moreover, because all of the edges radiated from a common focus, one place token might be assigned to the bunch of them; the dominant impression would be of oneness. Boundary tracing, (Ullman, 1984), the operation of moving the attentional focus along a contour, may be required to properly establish which edge belongs to which object. Of course, this laborious process could be cut short if the subject simply moved the attentional focus outward from the central fixation point and counted edge crossings. In any case, subjects would need to move the attentional focus in order to enumerate concentric objects. Consequently, subitizing concentric items should be difficult.

## Method

*Subjects.* Twelve undergraduate psychology students (5 men, 7 women) participated in the study for course credit.

*Apparatus and stimuli.* An Apple II+ computer was used to generate the displays and record the data. Vocal response time was measured using a Gerbrands G1341 voice-activated relay.

Displays were composed of up to eight white outline rectangles on a black background. Item positions and sizes were chosen randomly for each subject in each condition. In the same-size condition, all the rectangles were of the same size, although there were three possible sizes: $0.26° \times 0.16°$, $0.60° \times 0.42°$, or $1.01° \times 0.78°$ visual angle when subjects were seated 110 cm from the display. Rectangles could be located in any of 24 positions in a matrix. Assuming rectangles of the largest size, the closest horizontal and vertical neighbors were 1.2° and 0.94° from each other. Diagonal neighbours could be as close as 0.18°, however. At most, the entire display would occupy $8.02° \times 5.97°$ visual angle.

The different-size condition was the same except that at least one of the rectangles in the display was different in size from the others. Item sizes and positions were chosen randomly for each display and subject.

In the concentric condition, rectangles came in 15 sizes, ranging from $0.26° \times 0.16°$ to $7.25° \times 5.71°$ visual angle. For the inner six rings the minimal distance between items was 0.21° horizontal and 0.16° vertical. For the outer rings the distance was made larger because acuity decreases toward the periphery. Thus, for the outer rings the minimal distance was 0.29° and 0.21°, respectively. The maximal distance between rings was 3.49° horizontal and 2.71° vertical.

*Procedure.* Subjects were required to report the total number of rectangles in each display as fast as they could, with accuracy. Each trial had four phases. First, subjects were required to fixate on the central area of a white screen for 608 ms. The computer then beeped to indicate the start of the trial. The enumeration display came on 256 ms later. The display remained on the screen until the timer was activated. A white mask was then presented for 512 ms. (In this and all the following studies, postdisplay masks were used to prevent subjects from taking advantage of afterimages or phosphor decay to revise their number estimate after the display disappeared). Finally, the subjects were prompted to type in the number they had said or an "X" if something had gone wrong in the trial. The "X" response was reserved for situations in which the timer failed to go off the first time a vocal response was made, or went off before the response was made. These misfire trials were readministered at the end of each block.

There were 240 experimental trials preceded by 24 practice trials.

## Results and Discussion

The data are presented in Figure 4. Latencies were excluded from analysis if the subject's typed number response for a trial did not match the number of items in the display. Accuracy analyses will not be described because errors occurred through mistakes in typing as well as mistakes in counting. (Error rates were extremely low in the five experiments presented in this article: 1.9%, 1.4%, 4.5%, 1.8% and 2.4%, respectively. There was no evidence of speed–accuracy trade-offs).

All main effects and interactions were significant ($p < .05$). The effect of display was significant from two items on, $F(2, 22) = 18.4, p < .001$, at 2. Newman-Keuls analysis revealed that latencies for the concentric condition were greater than the other two conditions when subjects are enumerating 2 ($p < .05$), although there was no significant difference between same- and different-size latencies.

Trend analysis was performed on the entire range (1–8) to find out if subitizing occurred. Subitizing was clearly evident for multiple contour objects in the same- and different-size conditions. All subjects had nonlinear trends in the same- and different-size conditions. In contrast, there was little evidence of subitizing in the concentric condition. Only 2 of 12 subjects showed significant deviations from linearity in their latencies. Thus, a significantly greater proportion of the subjects showed nonlinear deviations in the same- and different-size conditions than in the concentric condition: $\chi^2(2) = 27.7, p < .01$.

Trend analyses were also performed on data sets made up of the average enumeration latencies for each subject. There were significant deviations from linearity in the same- and different-size conditions, $F(6, 77) = 12.2, p < .001$; $F(6, 77) = 13.5, p < .001$, consistent with the pattern of the individual data. Nonetheless, there was also a significant deviation in the concentric condition, $F(6, 77) = 2.3, p < .05$, though 10 of 12 subjects did not show deviations in their individual data.

Finally, the regression analyses also support the idea that subitizing was possible only in the same- and different-size conditions; slopes for the 1–3 range were 56 and 68 ms/item, respectively. The slope in the 1–3 range for the concentric condition was significantly greater, 212 ms/item, in contrast ($p < .05$). Slopes for the 5–7 range fell within 20 ms of each other in the three conditions ($p > .1$). For the same-size condition, the 5–7 slope was 306 ms/item; for the different-size condition, the 5–7 slope was 326 ms/item. The concentric condition fell intermediate at 310 ms/item in the 5–7 range.

There were discontinuities in slope between small and large numbers of items when subjects were enumerating multiple contour rectangles that were distributed across the screen. When subjects were required to enumerate concentric rectangles, however, the slope of the reaction time function was fairly constant, suggesting that the same processes were being used for small and large numbers of concentric rectangles. Although the concentric condition displays were symmetric, and the items did not necessarily fall into symmetric configurations in the same- and different-size displays, it seems unlikely that symmetry could explain the difference between conditions. Display symmetry does not seem to prevent subitizing (e.g., Frick, 1987; Mozer, 1989), or affect latencies when there are fewer than five items (e.g., Akin & Chase, 1978; Howe & Jung, 1987). The difficulty in the concentric condition would seem to stem either from the items being one inside another or sharing a common focus.

Two problems arose because of attempts to keep the display foveal, however. First, it was impossible to have as many different item sizes in the different-size condition as in the concentric condition without having items overlap. It is possible that subjects have great difficulty enumerating items of different sizes, but the different-size condition tested too-small a range for this difficulty to become apparent.

Second, on average, items were closer together in the concentric condition than the other conditions. Lateral interactions between edges might impede edge detection. Some take the precaution of ensuring that no item is within 0.5° of another to prevent lateral interference between items (e.g., Liss & Reeves, 1983; Sagi & Julesz, 1984), though Atkinson et al. (1976) showed that for short exposures subitizing accuracy does not begin to suffer until items are closer than 0.1°.

There was little evidence that the close proximity of contours from the same object interfered with enumeration, however. Recall that there were three item sizes in the same-size condition. For small rectangles, horizontal edges were 0.16° apart, and vertical, 0.26°. For large rectangles horizontal edges were 0.78° apart and vertical 1.01°. If simply having contours nearby slowed the process of edge extraction, presumably subjects would be slower at enumerating small rectangles than large because the contours are closer together, closer even than the suggested 0.5° limit. When analysis of variance was performed, comparing latencies for small, medium, and large rectangles, there were no significant differences, $F(2, 18) = 0.4, p > .1$. In fact, the small rectangles were enumerated 11 ms faster than large. Consequently, it seems that nearby contours do not impede enumeration if the contours come from the same item.

There is also interference from contours from different rectangles to consider, however. For this reason, a control study was performed. Subjects were required to enumerate nonconcentric items that were as crowded as the corners and sides of the concentric rectangles. In fact, in the control study subjects were required to enumerate isolated corners and sides from the concentric rectangles used in Experiment 1.

## *Experiment 2*

### *Method*

*Subjects.* Six subjects (4 men and 2 women) participated in the study for payments of $10. All were graduate students or summer research assistants at the University of Western Ontario.

*Stimuli.* The spacing of the stimuli was identical to that of items in the concentric condition in the previous experiment. In this experiment, however, subjects were shown parallel lines corresponding to one side of the rectangles, or parallel right angles corresponding to the corners of the rectangles.

Horizontal parallel lines ranged in length from 0.26° to 7.25°, whereas vertical lines ranged from 0.16° to 5.71° visual angle. Each corner was a right angle formed by connecting a 0.42° vertical segment with a 0.42° horizontal segment. The positions of these corners varied with the corners of the smallest (most central) to the largest (most peripheral) rectangle in the concentric condition of the previous study.

As before, the positions of the items were determined randomly. The top, bottom, left, and right sides of the boxes were displayed equally often, as were the top right, top left, bottom right, and bottom left corners. There were up to nine parallel lines or corners in each display. Only latencies for up to eight items were analyzed, however. The nine-item displays were used as catch trials to attenuate the end effect, the drop in latencies at the end of the range found in enumeration studies that measure reaction time.

*Procedure.* At the beginning of the session subjects were given 20 practice trials. The experimental session required 40 min. There were 312 experimental trials.

### *Results and Discussion*

Average enumeration latencies for correct trials are presented in Figure 5. Analysis of variance indicated no significant effect of display or interaction of display type with number ($p > .1$). There was clear evidence of subitizing. When trend analysis was performed on individual data sets, there were significant deviations from linearity for all subjects in all conditions. Deviations from linearity were also evident when datasets were averaged across subjects, $F(6, 63) = 14.3, p < .001$; $F(6, 63) = 14.0, p < .001$, for the parallel lines and corners conditions, respectively.

The 1-3 range slopes are very similar, 54 ms/item and 56 ms/item for parallel lines and corners, respectively. Moreover, these slopes are almost identical to the 56 ms/item slope for subitizing uniformly sized boxes in Experiment 1. The slope for the 5–7 range is slightly higher for the parallel lines than parallel corners, 368 ms/item as opposed to 333 ms/item. In this study the slopes in the 5–7 range are slightly higher than slopes from Experiment 1, which varied between 300 and 326 ms/item ($p > .1$, however).

Subjects were capable of subitizing crowded parallel corners and crowded parallel lines that varied in size by a factor of 30. Nonetheless, subjects had difficulty subitizing concentric rectangles that were made up of these parallel lines and corners. This may be because preattentive processing does not deliver clusters of contours that each correspond to an item for concentric items, either because items share a common focus or because items are one inside another. Further experimentation is needed to determine which of the two factors produces the problem. For present purposes, it is enough to show that subitizing is difficult under conditions that would be expected to produce preattentive misgrouping of contours.

## Subitizing Items in a Background of Distractors

People routinely enumerate certain target items while ignoring a number of distractors. The difficulty of this task may be influenced by how hard it is to determine whether an item is a target or distractor, however. We predict that subitizing should be possible only if the targets and distractors differ by a property that can be derived without spatial attention. Ullman (1984) argues that determining spatial relations such as connected requires attention (see Jolicoeur, 1988). Consequently, subjects should be unable to subitize connected items among distractors. Similarly, Treisman (1988) argues that finding an O among Qs or a conjunction of features (e.g., a green vertical line among green horizontal and white verticals) requires spatial attention because neither Os nor conjunctions pop out in search. Thus, we predict that subjects should not be able to subitize Os in Qs or conjunctions among distractors. Attention would not be necessary to find a target item that differs by a feature from the distractors, however, for example, finding an O among Xs or a green item among white items (e.g., Treisman & Gelade, 1980). Consequently we predict that subjects should be able to subitize items in a field of distractors in these situations.

Neither Ullman nor Treisman's theories explain subitizing. Nor do they predict that the attentional requirements of the enumeration task (dictated by the type of distractors) would affect whether subitizing occurs. According to these theories, subitizing should either never occur, regardless of the type of distractor items, or always occur, regardless of the type of distractor items.

### *Experiment 3: Enumerating Connected Items*

Ullman (1984) argues that determining whether two items are connected requires moving the attentional focus down the connecting contour from one item to the next (boundary tracing). Jolicoeur (1988) showed that the time to make judgments about whether two points were connected by a contour increased as the contour distance between the points increased, as would be expected if the attentional focus were being moved along a contour at a fixed rate. Thus, we predict that subitizing should be impossible when the task is to enumerate only the items connected by a specific contour when there are distractor items present. In contrast, enumerating items of a particular color in the presence of distractors should not require spatial attention. Subitizing should therefore be evident.

#### *Method*

*Design.* There were three factors in the study. The first was task. Subjects were required either to enumerate items of a specified color (color condition) or to enumerate items connected by particular lines (connected condition). Task was blocked so that subjects performed two color sessions and two connected sessions. Subjects did one session of each type before they did the second session of each condition. Otherwise session order was counterbalanced. The second factor was link length, the area in which the target items could appear. There were three possible link lengths, four, five, and six links, which corresponded to dense, medium, and sparse dispersion of targets in the display. The link-length factor was included as a manipulation check to ensure that subjects were in fact moving their attentional focus along the contour when enumerating connected items. The third factor was the number of target items that the subjects had to enumerate, between 1 and 8.

*Subjects.* Eight subjects (5 women and 3 men) participated in the four-session study for payments of $40. All were either graduate students or research assistants at the University of Western Ontario.

*Stimuli.* Each trial involved three displays: a fixation display, an enumeration display, and a mask. The fixation display was a 0.25° white square with a black asterisk inside, projected on a black background. The fixation point was used to direct subjects to where they were to start enumerating from. It could appear in three locations, the top left, top right, or bottom left corners of the display. In the connected condition, the fixation point would direct subjects to the end of the contour that they were supposed to trace. For example, if subjects were required to trace a horizontal contour the end of the contour might be in the top right corner. In order to keep the conditions equivalent, it was also necessary to direct subjects to the corners of the display in the color condition. For the color condition, the fixation simply indicated where to start enumerating from.

The enumeration display was made up of colored blocks and vertical and horizontal white lines projected on a black background. Subjects were required to enumerate square green or purple blocks. The sides of each block measured 0.25° visual angle. There were up to nine target blocks in each display, and either two, five, six, or eight distractor blocks. Only latencies for enumerating up to eight targets with six distractors were analyzed, however. The trials with nine targets or two, five, or eight distractors were included as catch trials to discourage guessing on the basis of simple contour density or the proportion of the field covered by a particular color. In the connected conditions the color of each target and distractor was determined randomly for each trial and subject.

Item locations were chosen from 84 positions on a matrix. The matrix was 5.7° × 6.74° visual angle. The blocks were located on white vertical and horizontal lines that were formed into a winding chain and a grid of orthogonal distractor lines. There were 12 different types of chain, as shown in Figure 6. The link length of the chain was simply the number of bends, or corners in the chain before there was a break that left segments unconnected with the rest of the chain.

Link length entailed slightly different things in the color and connected conditions, however. In the color condition, link length simply specified the area in which target items could appear. For example, for a four-link horizontal chain, target items would appear within a 6.74° × 3.8° area. For a similar four-link vertical chain, items could appear anywhere within a 5.7° × 4.5° area. For a five-link horizontal chain, target items could appear in a 6.74° × 4.7° area, whereas for a five-link vertical chain, items could appear in a 5.7° × 5.6° area. For six-link chains, the full 6.74° × 5.7° was used. In the connected condition, link length also specified the length of the contour that had to be scanned in order to enumerate all of the connected items, because of the requirement that target items be connected by the same line. For example, for a four-link horizontal chain, the maximum distance that would need to be scanned to find all the targets would be 37.5° visual angle. For a similar five-link chain, the distance would be 45.2°, and for a six-link chain, 52.9°. (Vertical chains were slightly shorter: 33.0°, 39.8°, and 46.6°, respectively.)

The procedure for positioning distractors also varied with condition. For the color condition, the distractors could appear anywhere in the grid except for where the targets were located. In Figure 6 there are five dark items in the sample display, for example. For the connected conditions, distractors could appear anywhere but on the designated chain before the first break. For example, there are six target items on the horizontal chain, enumerating from the top right corner in the sample display.

The masking display was a screen of colored random dots.

*Procedure.* The subject's task was to say how many blocks of a certain type there were in each display. Each trial had five phases. First, a message appeared on the screen informing the subjects which blocks they had to count. In the color sessions, the message said either "Count the green items" or "Count the purple items." In the connected sessions, the message said either "Count items connected by the vertical chain" or "Count items connected by the horizontal chain." These messages remained on the screen until subjects pushed the return key to initiate the trial. Next, the fixation point appeared. The fixation was used to inform the subject which corner to start enumerating from. The fixation point remained on for 2 s, and then the computer beeped to warn the subject that the trial was imminent. The enumeration display

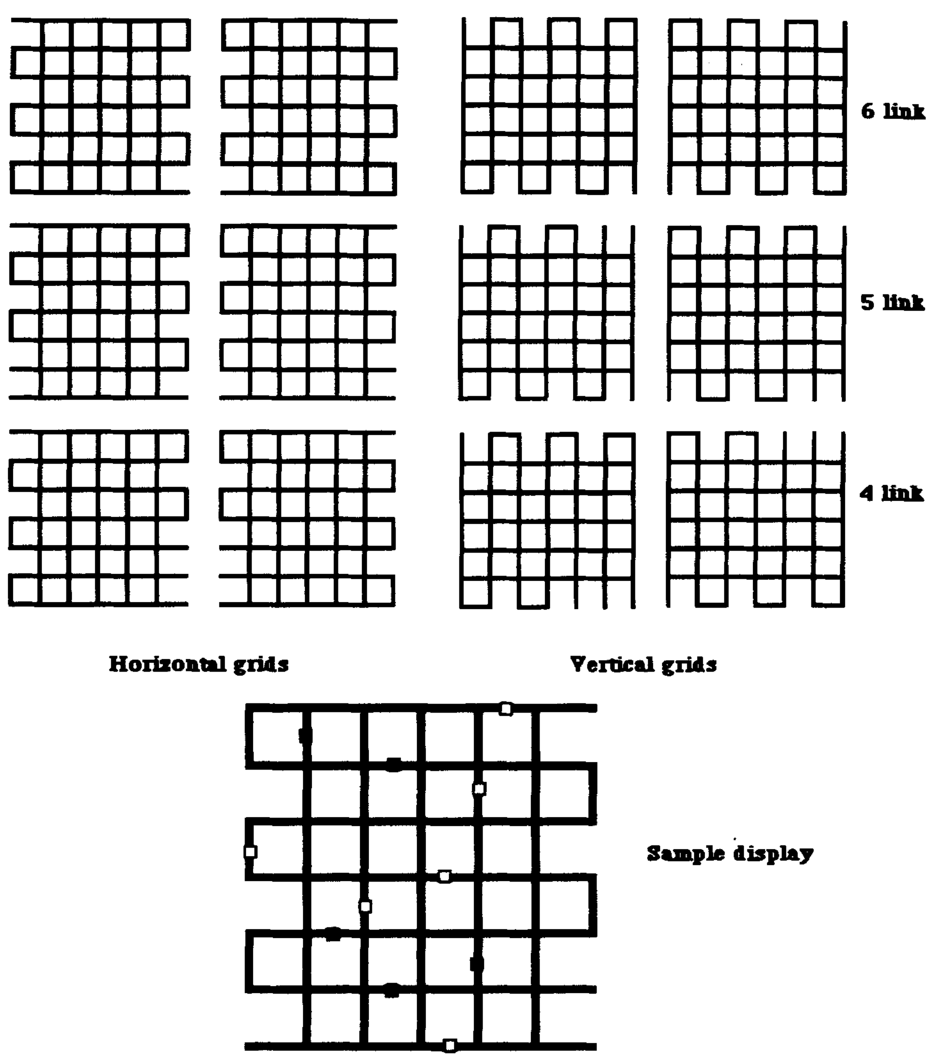


*Figure 6.* Examples of 12 types of winding contour and sample display for Experiment 3.

appeared 256 ms after this warning tone and remained until subjects made a response that set off the voice-activated timer. Then a colored random dot mask was presented for 512 ms. Finally, a message appeared on the screen asking subjects to type in the number that they had said or an "M" for mistrial. Error feedback was given.

There were 224 trials in each of the four sessions. Of these, 192 were experimental trials involving six distractors and between one and eight targets. The remaining 32 were catch trials in which there were nine targets or two, five, or eight distractors. Before the beginning of each session there were 28 practice trials. The experiment was run in 4 days.

### *Results and Discussion*

*Manipulation check.* If subjects were moving their attentional focus along the contour in the connected condition in order to determine how many connected items there are, link length should have had a strong effect on enumeration latency, as Ullman (1984) and Jolicoeur (1988) might predict. If subjects were simply moving their attentional focus to the targets in the color condition, the area that the targets occupy should have less effect, given that cuing studies show that the time required to move the attentional focus from location to location does not seem to reflect the distance between the locations (see Remington & Pierce, 1984). Average enumeration latencies are presented in Figure 7. As predicted, link length and condition interacted, $F(2, 14) = 62.3$, $p < .001$. Link length had a strong effect on latencies in the connected condition, $F(2, 14) = 99.2$, $p < .001$, with longer contours taking longer to process, as would be expected if subjects were boundary tracing to enumerate connected items. In contrast, when color condition trials alone were analyzed, link length, the area in

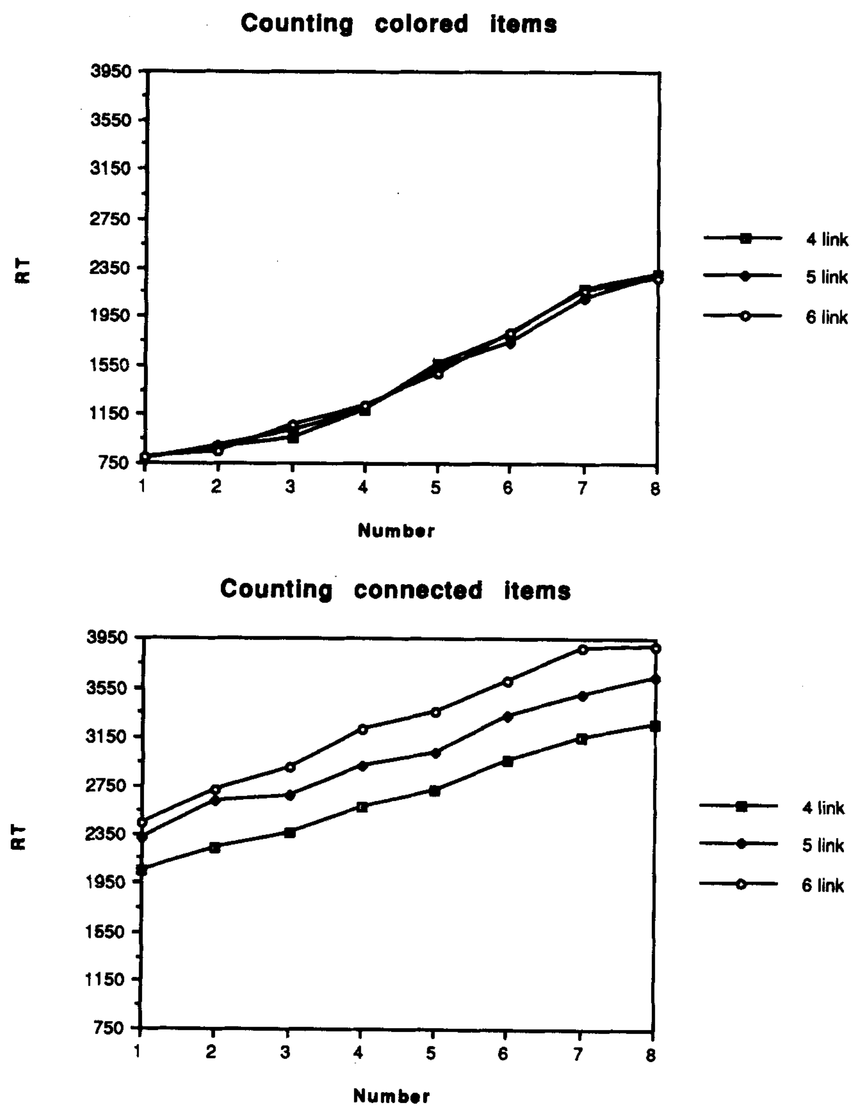


*Figure 7.* Average latencies to enumerate colored or connected items. (Standard errors for the means varied between 62.7 and 188.8, with larger standard errors associated with larger enumeration latencies. RT denotes reaction time.)

which the targets were concentrated, had no significant effect, $F(2, 14) = 0.9, p > .1$.

*Main analyses.* Subitizing was evident in the color condition even though displays were crowded with irrelevant lines and distractor items. All 8 subjects showed appropriate deviations from linearity in the color condition, regardless of the density of the target items (link length) or color of the target items. In contrast, there was little evidence of subitizing in the connected condition. There was only one case in which a deviation occurred because of a sudden increase in slope in the connected condition. Thus, there were significant differences in the proportions of subjects showing deviations from linearity in the color condition and the connected condition: $\chi^2(1) = 16.0, p < .01$, for the four- and six-link trials; $\chi^2(1) = 12.04$, for the five-link trials.

Trend analysis on averaged data sets revealed a similar pattern of results. When subjects were required to enumerate items of a particular color, significant deviations from linearity occurred, regardless of the color of the targets or the density of the items, $F(6, 49) = 6.5, 9.3, 5.0$, for the 4–6–link green items ($p < .01$), and $F(6, 49) = 5.3, 3.6$, and 3.8, for the 4–6–link purple items ($p < .01$). In contrast, when subjects were required to enumerate connected items, there was only a significant deviation from linearity in the six-link condition, $F(6, 49) = 2.9, p < .05$, and that deviation was the result of a slight drop in latency at 8, $F(5, 49) = 0.0003$ (if only the 1–7 range is considered). Otherwise there were no significant deviations from linearity in the connected condition, $F(6, 49) = 0.47$; $F(6, 49) = 1.64$, for four- and five-link trials, $p > .1$.

Finally, regression showed that for the color condition, the slopes in the 1–3 range were 86, 99, and 135 ms/item for 4–6–link areas, respectively. The corresponding slopes in the 5–7 range were 300, 295, and 353 ms/item for the 4–6–link areas.

For the connected condition, the slopes in the 1–3 range were 170, 191, and 239 ms/item for the 4–6–link chains, respectively. The corresponding slopes in the 5–7 range were lower than the 5–7 range slopes in the color condition and similar to the connected condition in the 1–3 range. The slopes were 239, 233, and 233 ms/item for the 4–6–link chains. The reduced slope in the 5–7 range for the connected condition undoubtedly contributed to the finding that there were no significant deviations from linearity in the connected condition. Nonetheless, it seems as if 1–3 items were processed differently from 5–7 items from the difference in slopes in the color condition, whereas it seems as if 1–3 and 5–7 items were processed similarly in the connected condition.

There was one surprising finding: the effect of the particular color of the target item on the slopes in the color condition. When the data for green and purple items were analyzed separately, it became apparent that the 1–3 range slope for purple items was almost twice that of green, 141 ms/item as opposed to 77 ms/item ($p < .05$). In contrast, in the 5–7 range slopes for green and purple were almost identical, 310 ms/item as compared with 305 ms/item, respectively ($p > .1$). Thus, it appears that either the hue, saturation, or brightness of the items to be enumerated had an effect on the subitizing slope but not on the counting slope. A pilot study revealed that subjects require approximately 55 ms longer to locate a purple item in a field of green and white distractors than to locate a green item in a field of purple and white items in search. Nonetheless, the search slopes were comparable: 2.8 ms/item and 2.9 ms/item for positive and negative trials for green items, and 3.8 ms/item and 4.9 ms/item for purple items. Why did the interaction between color and number occur? In particular, why did target color affect enumeration in the subitizing range and not the counting range? It seems unlikely that subjects were moving attentional focus to enumerate up to three items in the purple condition because there was a consistent slope discontinuity at every link length for every 1 of the 8 subjects for the purple items.

Perhaps the contrast between target and background produced the interaction. Purple items contrasted less with the background than green.[3] Two previous studies suggest that there is a relationship between the subitizing slope and contrast. Hunter and Sigler (1940) showed that there was a relationship among contrast, number, and the required exposure duration for 50% enumeration accuracy for small but not large numbers of items.[4] More recently, in a reaction time study, Liss and Reeves (1983) showed that decreasing the exposure duration to 3 ms (without masking) caused the 1–3 slope to increase markedly though the 4–7 slope dropped a small amount. (Decreasing the exposure duration would effectively decrease the contrast between the target items and the background.) The finding that contrast-related factors affect the subitizing slope more than the counting slope may indicate that subitizing is data limited and counting is resource limited (Norman & Bobrow, 1975). Subitizing would be data limited because subitizing occurs when there are more FINSTs than items, and thus the principal limitation is stimulus quality. Counting would be less affected by stimulus quality because the principal limitation in counting is access to the attentional focus resource.

In this study, however, the relationship between the color and the subitizing slope might reflect a greater tendency to perform eye movement in purple target trials. Eye movements have been shown to inflate subitizing slopes (Klahr, 1973b). If it were difficult to discriminate purple items from the background, subjects might routinely perform eye movements for trials with purple targets. In contrast, subjects might only perform eye movements for green target trials when the targets were far away from

[3] The illuminance in the room was 135 $mW/m^2$, as measured by a Techtronix J6502 photometer. A single green item radiated 55 $cd/m^2$ at a 1-m distance. A single purple item radiated only 41 $cd/m^2$, a value similar to the luminance of an area containing a white contour line on a black background—38 $cd/m^2$. (The latter measurements were made using a Techtronix photometer fitted with a J6523 1° narrow luminance probe adjusted to match the spectral sensitivity function.)

[4] In their study "small" was less than 8, whereas in the present "small" is less than 4, though Schlosberg (1948) commented that the relationship between contrast and duration began to break down after 4 in Hunter and Sigler's study.

each other, which would be more probable when there were large numbers of items.

## *Experiment 4: Enumerating Os in Distractor Letters*

We predicted that subitizing would be possible only in cases where targets pop out among distractors in search, in cases where preattentive information was adequate to distinguish targets from distractors. Os pop out among Xs; an O can be detected in approximately the same time, regardless of the number of Xs in the display. In contrast, the time required to locate an O in a field of Qs varied with the number of targets in the display, as would be expected if it were necessary to move the attentional focus from item to item to locate an O among Qs (Treisman, 1988). Consequently, we predicted that subjects would be able to subitize Os among Xs but not Os among Qs.

### *Method*

*Design.* There were three factors. The first was condition. In the OX condition, subjects enumerated Os in X distractors. In the OQ condition, subjects enumerated Os in Q distractors. The second factor was the number of distractors. There were either zero, two, or four distractor letters. (The number of distractors was chosen so that the classical pattern for attentive search, the 2:1 ratio for negative to positive slopes, would obtain. A pilot study revealed that this ratio occurred even when subjects were searching for an O among only two and four Qs. The ratio of negative to positive slopes was approximately 2:1, 208 ms/item to 117 ms item). The third factor was number. Subjects were required to enumerate between one and eight Os.

The condition factor was blocked so that subjects did the OX session on 1 day and the OQ on another. Session order was counterbalanced. All other factors were randomized.

*Subjects.* Twelve subjects (8 women and 4 men) participated for payments of $20. The subjects were graduate students and research assistants from the University of Western Ontario.

*Stimuli.* The displays contained up to 15 white letters from the standard Apple character set. The background color was black. There were in most cases 1–8 Os and either 0, 2, or 4 distractor letters. Forty catch trials were also included in which there were no Os or nine Os, or one, three, or seven distractor letters to prevent subjects from guessing on the basis of simple contour density.

Letters could occupy any position in a 49-point matrix, which occupied $5.97° \times 4.2°$. Each letter occupied $0.36° \times 0.21°$ visual angle. The minimal horizontal distance between items was 0.73°, and the minimal vertical distance was 0.36°. Target and distractor positions were chosen randomly for each trial and subject.

A white mask was presented after the enumeration display. The screen had a 0.25° fixation cross affixed.

*Procedure.* Each trial had four stages. First, during the 512-ms pretrial interval, the screen went white, and subjects were required to fixate on a green fixation cross. The computer then beeped to signal the start of the trial. After 256 ms, the letter display appeared and remained on until subjects made a vocal response. Third, as soon as the timer registered the response, the display disappeared, and the screen turned white. Finally, after 512 ms subjects were prompted to type in the number that they had said, or an "M" for mistrial.

The experiment was run in two sessions of 328 trials. At the beginning of each session subjects were given 60 practice trials.

### *Results and Discussion*

*Manipulation check.* If subjects have to move their attentional focus from item to item in the OQ condition but not the OX condition, as Treisman (1988) suggests, the number of distractors should have a stronger effect on latencies in the OQ condition than the OX condition. As can be seen from Figure 8, this prediction was borne out. The number of distractors had a much stronger effect in the OQ condition than the OX condition, $F(2, 22) = 1{,}071.8$, $p < .001$, as is typical in search studies of this type (Treisman, 1988). Nonetheless, the number of distractors had a significant effect in both conditions ($p < .05$). Each distractor added on average 36 ms to the times to enumerate 1 in the OX condition, which may reflect the effects of lateral inhibition between items, eye movements, or the cost of filtering (Kahneman & Treisman, 1984; Treisman, Kahneman, & Burkell, 1983). Each distractor added on average 280 ms to the latency to enumerate 1 in the OQ condition, which might be thought of as the cost of feature integration.[5]

*Main analyses.* Subjects only seem to be capable of subitizing Os in a field of distractors when the Os and distractors differed on the basis of a feature that produced pop out. Although most subjects had significant deviations from linearity when there were no distractors in the OX and OQ conditions, (12/12 and 10/12 respectively), there was clear evidence of the ability to subitize targets items in the OX condition only. None of the subjects had deviations in the OQ condition (0/12 in both the two- and four-distractor conditions), whereas most had deviations in the OX condition (11/12 and 9/12, respectively). Thus, significantly more subjects had deviations from linearity in the OX condition than in the OQ condition when there were distractors: $\chi^2(1) = 20.3$, $p < .01$; $\chi^2(1) = 14.4$, $p < .01$, for the two- and four-distractor cases, respectively.

The results were similar when trend analyses were performed on averaged data sets. There were significant deviations from linearity in both the OX and OQ conditions when there were no distractors, $F(6, 77) = 20.6$, $p < .01$; $F(6, 77) = 11.01$, $p < .01$, respectively. Once there were distractors, there were no significant deviations in the OQ condition, $F(6, 77) = 1.1$, $p > .1$, and $F(6, 77) = 1.7$, $p > .1$, for two and four Q conditions, respectively, though there were significant deviations in the OX condition, $F(6, 77) = 6.9$, $p < .01$, and $F(6, 77) = 4.6$, $p < .01$, for two and four X conditions, respectively.

Regression revealed a steady increase in the 1–3 range slope in the OX condition as the number of distractors increased, from 44 to 94 to 108 ms/item, though the addition of the first two Xs had the greatest effect, more than dou-

[5] The corresponding slopes in the search pilots were less than half as large, however: 8 ms in the OX condition and 117 ms /item in the OQ condition.

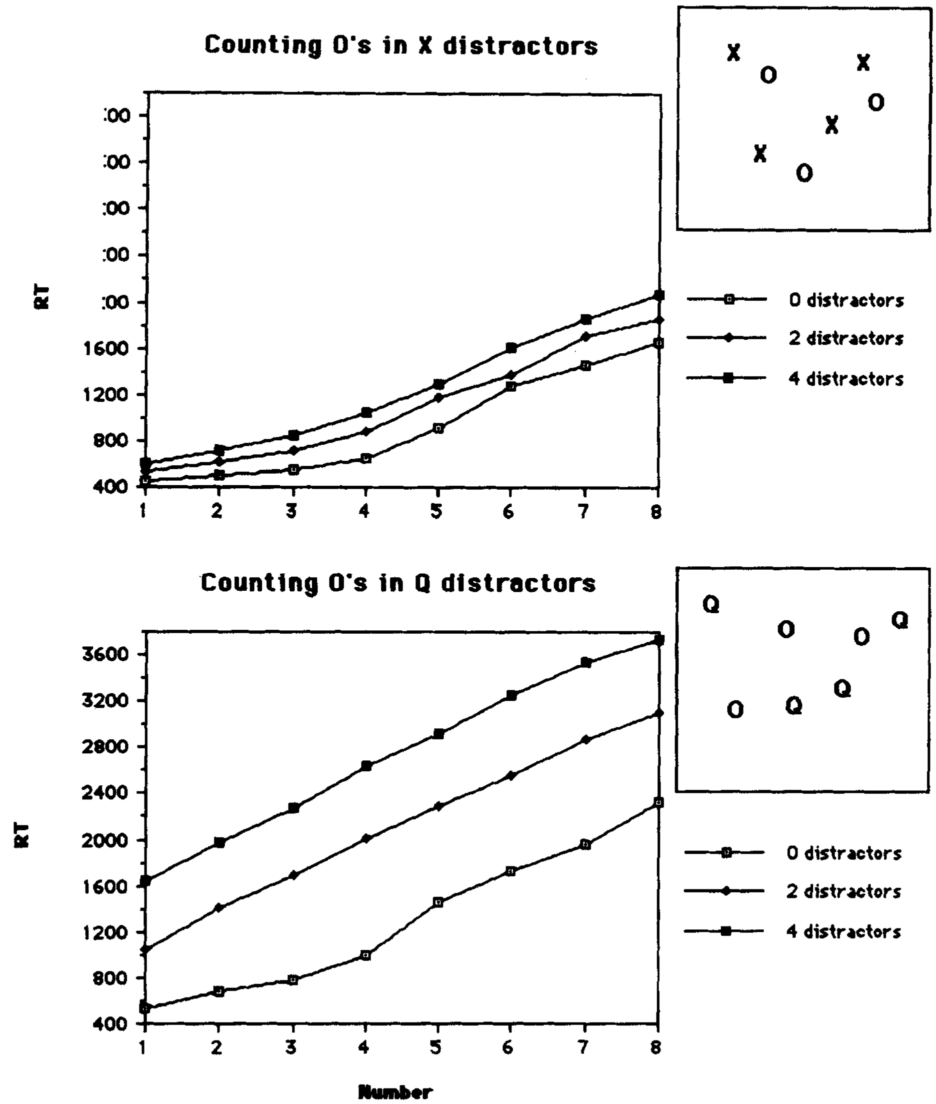


*Figure 8.* Average latencies to enumerate Os in a background of X or Q distractors. (Standard errors for the means varied between 13.3 and 92.4, with larger standard errors associated with larger enumeration latencies. RT denotes reaction time.)

bling the slope. The corresponding slopes in the 5–7 range were 276, 262, and 286 ms/item.

In contrast, in the OQ condition the slopes in the 1–3 range were considerably larger. The addition of the first two Qs increased the slope from 129 ms/item to 323 ms/item in the OQ condition. The addition of the second pair of Qs decreased the 1–3 range slope to 314 ms/item, however. In the 5–7 range, the slopes were 255, 295, and 310 ms/item in the zero-, two-, and four-distractor conditions.

There were significant differences between OX and OQ conditions in the 1–3 range slopes even when there were no distractors, however. The subitizing slope was significantly greater in the OQ condition, 129 ms/item as compared with 44 ms/item ($p < .05$). Perhaps subjects were being especially cautious in the OQ condition even when there were no distractors present.

Although the present results are compatible with the idea that subitizing is not possible when attentional processing is required to combine parts, there is an different interpretation of the data. Distinguishing Os from Qs presumably requires more high-resolution analysis than distinguishing Os from Xs. Therefore, all that can be concluded is that in situations where the targets and distractors are very similar, serial processing is required, and consequently, subitizing is not observed (see Duncan & Humphreys, 1989). For this reason, another experiment was performed. In it, stimuli varied on two dimensions, color and orientation, and Treisman's conjunction methodology was used. If the present results replicate with colored lines, then it seems less likely that the only reason that subjects could not subitize in the OQ condition was the need for high-resolution analysis.

### *Experiment 5: Enumerating Disjunctions and Conjunctions of Color and Orientation*

#### *Method*

*Design.* There were three factors. The first was task. In the conjunction condition, subjects were required to enumerate white vertical lines in a field of green vertical and white horizontal lines. In the disjunction condition, subjects were either required to enumerate white horizontal lines in a field of green horizontals or green vertical lines in a field of green horizontals. The second factor was the number of distractors in the display; there were either 0, 12, or 20 distractors. The number of distractors was chosen to produce the classic pattern of results for conjunction search, the 2:1 ratio of negative to positive slopes. In our preliminary search study we found the the the ratio of negative to positive slopes did not approach 2:1 until until there were more than eight items, as found by Pashler (1987).[6] The third factor was the number of targets, between one and eight.

Subjects performed the conjunction session on 1 day and the disjunction sessions on the other 2 days. Half of the subjects started with the conjunction session, and half started with the disjunction sessions.

*Subjects.* Ten subjects (5 women, 5 men) participated in the three-session study for payments of $30. Subjects were graduate students or research assistants from the University of Western Ontario.

*Stimuli.* Each trial involved three displays: a fixation display, an enumeration display, and a mask. The fixation display was a colored random dot mask with a 0.26° black square in the center. The final display, the mask, was a colored random dot display projected on the screen after the enumeration display disappeared.

The enumeration displays were composed of 1–29 line segments. These line segments could be vertical or horizontal, green or white. Each segment was 0.26° visual angle. In each display there could be 0–9 target items and 1, 3, 7, 12, or 20 distractors. Trials with 0 or 9 targets or 1, 3, or 7 distractors were used as catch trials. The appearance of the target and distractor items was dictated by condition. In the disjunction condition trials, targets were either white horizontal lines or green vertical lines, and distractors were green horizontal lines. In conjunction trials, the targets were white vertical lines, and the distractors were white horizontal or green vertical lines. There were approximately equal numbers of white horizontal and green vertical lines when there were distractors in the conjunction condition.

The target and distractor items could fall on any of 40 locations in a matrix. The entire matrix occupied 4.42° × 5.19° visual angle. The minimum horizontal distance between items was 0.78°, whereas the minimal vertical distance was 0.94°. Stimuli could fall within 0.63° if they were diagonal neighbors, however. The positions of the target and distractor items were chosen randomly for each trial and subject.

*Procedure.* The subject's task was to enumerate specified line segments as fast as possible, with accuracy. Each trial had four phases. First, the fixation display was shown for 512 ms. During this time the computer beeped to warn the subject that the trial was about to start. The enumeration display appeared 256 ms later and remained until the subject set off the timer. Next, a colored mask was projected for 512 ms. Finally, the computer prompted the subject to type in the number that they had said, or an "M" for mistrial if something had gone wrong during the trial. If the subjects typed in the wrong number, the computer beeped five times and reminded them what kind of items they were supposed to count.

Each session had 328 trials, 288 experimental trials and 40 catch trials. At the beginning of each session subjects did 36 practice trials.

#### *Results and Discussion*

Subjects did two sessions of the disjunction condition in order to guarantee that there would be enough observations so that vertical and white item latencies could be compared. Whenever comparisons are made between conjunction and disjunction conditions, however, only the first session disjunction trials are considered. This ensures that subjects are no more practiced in the disjunction condition than in the conjunction condition trials.

*Manipulation check.* If subjects needed to move their attentional focus from item to item in the conjunction condition but not in the disjunction condition, as Treisman and Gelade (1980) suggest, the number of distractors should have a stronger effect on latencies in the conjunction condition than in the disjunction condition. The average enumeration latencies for the two conditions are displayed in Figure 9. The number of distractor items had an effect on latencies in both the disjunction and conjunction conditions ($p < .01$), though the effect was stronger in the conjunction condition, $F(2, 18) = 69.6$, $p < .001$, for the interaction. In fact, in the disjunction condition the first 12 distractors added only 12.6 ms/distractor, and adding another 8 distractors (to make 20) subtracted approximately 3.3 ms/item from this value. Thus, in the disjunction condition each distractor cost 6.2 ms on average. In the conjunction condition, the first 12 distractors contributed 67 ms/distractor, and the remaining 8 contributed an additional 61 ms/distractor, for an average of 65 ms/distractor.

*Main analyses.* There was clear evidence of subitizing target items in distractors only when the targets and distractors differed by a feature, as in the disjunction condition. Trend analyses revealed deviations from linearity in the latency of all subjects, regardless of task, when there were no distractors. Once there were distractors, however, greater proportions of subjects had deviations from linearity in the disjunction condition than in the conjunction condition, 9/10 as compared with 0/10 for the 12-distractor case, $\chi^2(1) = 16.3$, $p < .01$, and 9/10 as opposed to 1/10 in the 20-distractor case, $\chi^2(1) = 12.8$, $p < .01$.

Analyses on the averaged data set revealed a similar story. There were significant deviations from linearity when the task was enumerating white or vertical lines. Whether there were 0, 12, or 20 distractors, there was always a deviation from linearity in the disjunction condition: $F(6, 63) = 11.4$, $p < .001$; $F(6, 63) = 4.7$, $p < .001$; $F(6, 63) = 4.3$, $p < .001$, for the 0-, 12- and 20-distractor cases. In the conjunction condition there were significant deviations from linearity only when there were no distractors in the display, $F(6, 63) = 9.3$, $p < .01$, as compared with the 12-distractor trial, $F(6, 63) = 0.5$, and and the 20-distractor trial, $F(6, 63) = 0.4$.

[6] For Display Sizes 13–21, the slope for negative trials was 34.3 ms/item, and the slope for positive trials was 19.6 ms/item.

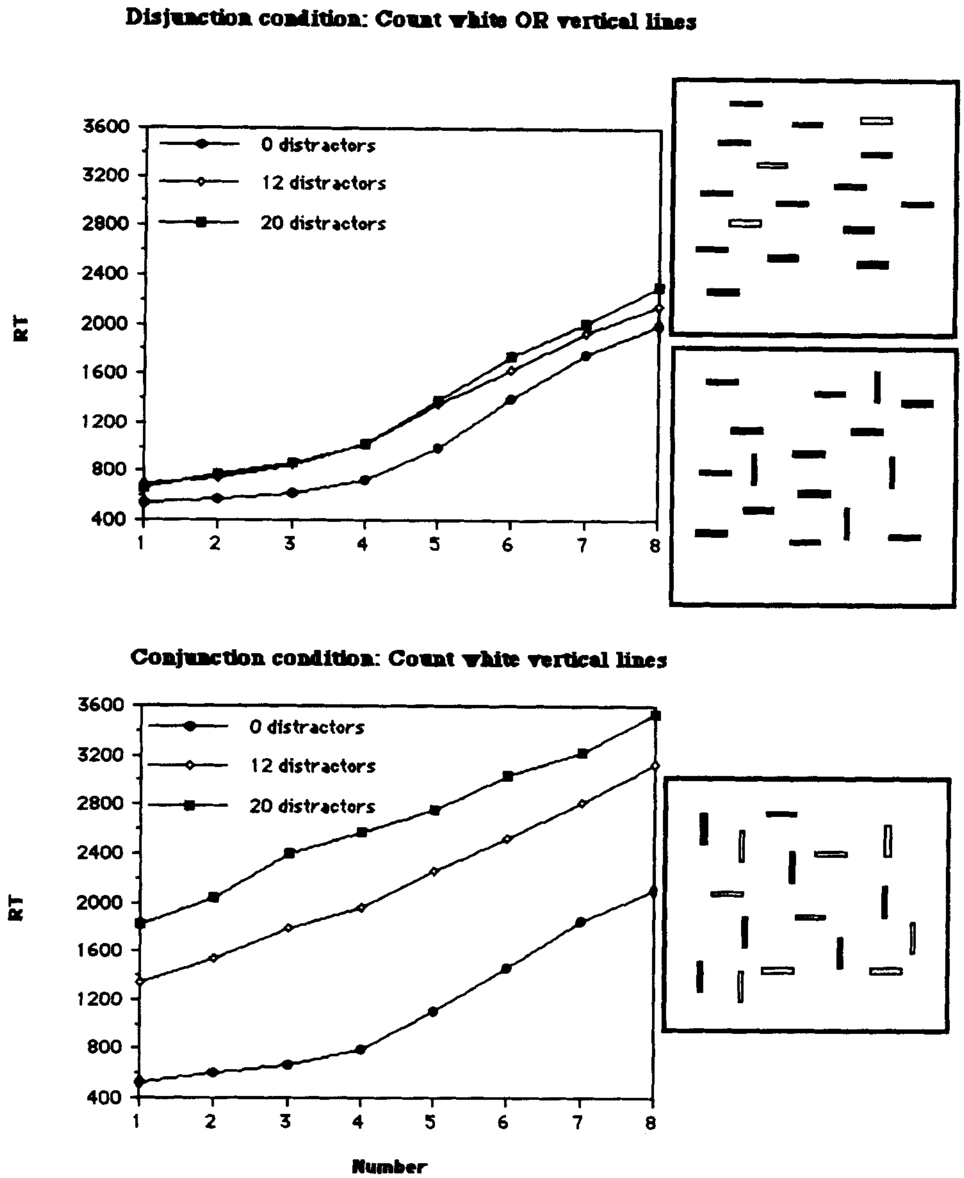


*Figure 9.* Average latencies to enumerate white vertical as opposed to white or vertical lines. (Standard errors for the means varied between 21.2 and 261.5, with larger standard errors associated with larger enumeration latencies. RT denotes reaction time.)

In the disjunction condition, the 1–3 slope gradually increased with the addition of distractors, from 49 to 65 to 92 ms/item. The corresponding slopes in the 5–7 range were 366, 283, and 326 ms/item for 0, 12, and 20 distractors.

In the conjunction condition, there was evidence of subitizing only when there were no distractors. The 1–3 range slope was 60 ms/item. (This slope is more than the average 49 ms/item observed with the same number of distractors in the disjunction condition and more than the slopes for either white items [49 ms/item] or vertical items [49 ms/item].) For 12 and 20 distractors, the 1–3 range slopes were 209 and 273 ms/item. In the 5–7 range, the slopes were 361, 266, and 225 ms/item in the 0-, 12-, and 20-distractor conditions.

The results of this study are consistent with the previous study in that both show that when attentional processing is required to distinguish targets from distractors, subitizing is not evident. Consequently, even when there is no need for high-resolution processing, subitizing disappears with the need for spatial attention.

Notice that the 1–3 slope for disjunction enumeration is far in excess of the usual slopes for conjunction search. For example, the subitizing slopes are 65 ms/item and 92 ms/item for 12 and 20 distractors for disjunction enumeration as opposed to Treisman and Sato's (1990) estimate of 11.7 ms/item for conjunction search for color and orientation, with display sizes of 4–16 items. Does this mean that the attentional focus is being moved from item to item even in disjunction enumeration?

Probably not. First, although the 1–3 range slopes are high in the disjunction condition, they are much lower than the 5–7 range slopes. There is nothing in anybody's theory of attention that suggests that when the attentional focus visits items in conjunction search, it suddenly needs 200 ms more per item after the fourth distractor. If it were true, in search there should be a discontinuity, a sudden increase in the slope after 4 distractors. Furthermore, if attention is being moved from item to item, why would there be a difference in slopes between disjunction and conjunction conditions in the 1–3 range but not the 5–7 range?

If the subitizing slope does not come from moving the attentional focus, where does it come from? A review of the enumeration literature suggests that subitizing slopes originate primarily from the need to choose a number response from a range of responses. For example, in Klahr's (1973a) production system, the subitizing slope originated partly from the need to associate individuated items with number names, in the order of the number names. Evidence supports this contention. The subitizing slope decreases markedly when the response choice part of the latency is excluded by using a threshold procedure, where stimulus onset asynchrony between stimulus and mask required for a certain level of accuracy is measured (e.g., Oyama et al., 1981, 4–10 ms/item). The subitizing slope all but disappears if this methodology is used, and the number of response alternatives is reduced to two (Sagi & Julesz, 1984, 1.9 ms/item). Therefore, our subitizing slope was as high as it was because we used reaction time as dependent measure, had a wide range of response alternatives (9 usually), and did not control eye movements, which have been also shown to inflate subitizing slopes (Klahr, 1973b).[7]

If comparisons between search and enumeration are to be made, it is more instructive to look at the time to enumerate one item given a varying number of distractors rather than looking at the subitizing slope, which is the time to enumerate varying numbers of targets given a set number of distractors. The time to enumerate one item is the equivalent to search latencies for trials in which the target is present in the display. Search pilots were performed with the same stimuli as the enumeration studies and exactly the same numbers of distractors (0–20). Generally, the slopes per distractor in enumeration studies were twice the slopes for the search pilots. Each distractor cost 6.2 ms when enumerating one item in disjunction enumeration but only 3.4 ms/item in disjunction search. Similarly, each distractor cost 65 ms in conjunction enumeration but only 32 ms in conjunction search. The doubling of the time per distractor may reflect the fact that in enumeration subjects are required to respond to the presence of all the targets, not just the first one they see. Moreover, we used aversive error messages (multiple computer beeps) to ensure that subjects enumerated with accuracy.

## General Discussion

These studies converge on a common conclusion. When there was no need for attentional processing, either because the property that distinguished targets from distractors was a preattentive feature, such as color or orientation, or because preattentive grouping processes delivered feature clusters that each corresponded to an item, subitizing was evident from latency data. Specifically, there were significant deviations from linearity in the latency functions, and the slopes in the 1–3 range were low. When attention was required to compute a spatial relation, to resolve the object as a whole, or to combine parts or features into an object description, there was little evidence of subitizing. There were no deviations from linearity in the latency functions, which suggests that similar processes were being used for both ranges. Furthermore, the slopes in the 1–3 range were high; it is the counting process that was being used for small and large numbers of items. These findings are consistent with the idea that subitizing can occur only when preattentive information is adequate to resolve items as wholes and distinguish the items to be enumerated from distractors.

In order for preattentive information to be useful for enumeration, there must be some way to individuate the feature clusters derived from preattentive analysis. There needs to be a way of individuating each feature cluster, so it can considered a separate unit. Pylyshyn's (1989) FINST mechanism provides a way to accomplish this, and moreover, this mechanism would be able to compensate for any moment-to-moment changes in the retinal location or retinal properties of the items to be enumerated.

These results conflict with predictions from Ullman's (1984) visual routines and Treisman and Gelade's (1980) feature integration theory. Visual routines were envisioned to explain enumeration as well the computation of spatial relations. Nonetheless, there is no mechanism in the theory to explain the slope discontinuity in the enumeration function indicative of subitizing. Feature integration theory posits a parallel preattentive stage that has no mechanism for the individuation of different items with the same features within a field of distractor items. Neither theory predicts that attentional manipulations would affect whether subitizing occurs. At this point, it is possible to re-address the questions set at the outset of the article. Why are small numbers of items enumerated differently from large numbers of items? Why can humans not subitize any number of items? According to the FINST hypothesis, the reason is that humans have a limited-capacity mechanism for individuating feature clusters. There are only a small number of spatial reference tokens or FINSTs. When the number of items exceeds the number of FINSTs a different process is required, the counting process proper.

The advantage of the FINST account is that it explains the flexibility of the subitizing process. Subjects can subitize multiple contour items of different sizes. They can subitize items in a field of distractors, as long as the targets and distractors differ by a feature. They can subitize even when items do not fall into their characteristic patterns. For ex-

[7] There are a number of other factors that influence the absolute slope in subitizing, however: the age of the subjects (Chi & Klahr, 1975; Svenson & Sjoberg, 1978), the complexity of the items (cf. Akin & Chase, 1978, with typical dot enumeration studies), and the discriminability of the items to be enumerated (Hunter & Sigler, 1940; Liss & Reeves, 1983), for example.

ample, humans subitize three dots even if the dots do not fall into a triangular configuration (e.g., Frick, 1987; Trick, 1987). Moreover, it seems that visual capacity rather than working memory capacity limits the number of items that can be subitized because subjects can subitize four different-size boxes but not four concentric boxes, four Os in four Xs but not four Os in four Qs. There is no reason to expect that concentric boxes or Os in Qs would demand more working memory.

What can this research tell us about vision in general? Consider the following points. First, the studies presented here suggest that subitizing relies on preattentive information to resolve items as wholes and to distinguish items to be enumerated from distractors. Second, subitizing must exploit some limited-capacity mechanism because it is possible to subitize only small numbers of items (around four or so). Third, subitizing must involve individuation because enumeration by its nature requires individuation; each item must be considered as an individual in order to ensure that items are not repeatedly enumerated or missed. (It is important to keep the item currently being assigned a number name distinct from items already assigned number names, and those yet to be assigned number names.) By putting these ideas together, we may surmise that there is a limited-capacity preattentive mechanism for item individuation.

This sort of mechanism would not only be useful for subitizing, however. In order for us to deliberately move our attentional focus from item to item in a display, there is a logical need for preattentive item individuation. It would be better to have this individuation mechanism rely on internal reference tokens, such as FINSTs, rather than the properties or retinal location of items, because properties and locations change from moment to moment. If we are to move our attentional focus toward a specific item, there needs to be a way to consider it the same one even if it moves and changes properties.

Moreover, there is a logical need to be able to individuate not just one, but several, if we are to use the attentional focus to compute spatial relations, as several researchers suggest (e.g., Treisman, 1988; Ullman, 1984). For example, in order to determine whether one item is connected to another by a contour, there needs to be a way to consider each item as a separate unit. Similarly, there is a need to individuate not just one, but several, to perform multiple target tracking (Pylyshyn & Storm, 1988). A preattentive mechanism is needed to explain how we can track four or five items in as many identical distractors when all are set into independent motion and the targets are traveling too fast to be monitored by an attentional scan from target to target.

Therefore, there must be a limited-capacity preattentive item individuation process that operates after feature detection and grouping but before spatial attention. This process represents a halfway point between parallel and serial analysis, a stage that is parallel and yet limited capacity.

## References


Akin, O., & Chase, W. (1978). Quantification of three-dimensional structures. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 397–410.

Aoki, T. (1977). On the counting process of patterned dots. *Tohoku Psychologica Folia, 36,* 15–22.

Atkinson, J., Campbell, F., & Francis, M. (1976). The magic number 4 + 0: A new look at visual numerosity judgements. *Perception, 5,* 327–334.

Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and representation in perception.* Hillsdale, NJ: Erlbaum.

Chi, M., & Klahr, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology, 19,* 434–439.

Duncan, J., & Humphreys, G. (1989). A resemblance theory of visual search. *Psychological Review, 96,* 433–458.

Egeth, H., Virzi, R., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 32–39.

Folk, C., Egeth, H., & Kwak, H. (1988). Subitizing: Direct apprehension or serial processing? *Perception & Psychophysics, 44,* 313–320.

Frick, R. (1987). The homogeneity effect in counting. *Perception & Psychophysics, 41,* 8–16.

Howe, E., & Jung, K. (1987). Judgement of numerosity: Effect of symmetry and goodness of dot patterns. *Acta Psychologica, 64,* 3–12.

Hunter, W., & Sigler, M. (1940). The span of visual discrimination as a function of time and intensity of stimulation. *Journal of Experimental Psychology, 26,* 160–179.

Jensen, E., Reese, E., & Reese, T. (1950). The subitizing and counting of visually presented fields of dots. *Journal of Psychology, 30,* 363–392.

Jevons, W. (1871). The power of numerical discrimination. *Nature, 3,* 281–282.

Jolicoeur, P. (1988). Curve tracing operation and the perception of spatial relations. In Z. Pylyshyn (Ed.), *Computational processes in human vision: An interdisciplinary approach* (pp. 133–168). Norwood, NJ: Ablex.

Julesz, B. (1984). Towards an axiomatic theory of preattentive vision. In G. Edelman, W. Gall, & W. Cowan (Eds.), *Dynamic aspects of neocortical function.* Toronto: Wiley.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. Davies (Eds.), *Varieties of attention* (pp. 29–61). Orlando, FL: Academic Press.

Kaufman, E., Lord, M., & Reese, T., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62,* 498–525.

Klahr, D. (1973a). A production system for counting, subitizing and adding. In W. G. Chase (Ed.), Visual information processing (pp 527–546). New York: Academic Press.

Klahr, D. (1973b). Quantification processes. In W. G. Chase (Ed.), *Visual information processing* (pp. 3–34). New York: Academic Press.

LaBerge, D., & Brown, V. (1987, November). *Comparison of moving-spotlight and gradient theories of attention.* Paper presented at the annual meeting of the Psychonomic Society, Seattle, WA.

Liss, P., & Reeves, A. (1983). Interruption of dot processing by a backward mask. *Perception, 12,* 513–529.

Mandler, G., & Shebo, B. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111,* 1–22.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Milligan, G., Wong, D., & Thompson, P. (1987). Robustness properties of nonorthogonal analysis of variance. *Psychological Bulletin, 101,* 464–470.

Mozer, M. (1989). Types and tokens in visual letter perception. *Journal of Experimental Psychology: Human Perception and*

*Performance, 15,* 287–303.
Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44–64.
Oyama, T., Kikuchi, T., & Ichihara, S. (1981). Span of attention, backward masking and reaction time. *Perception & Psychophysics, 29,* 106–112.
Parkman, J., & Groen, G. (1972). Temporal aspects of simple addition and comparison. *Journal of Experimental Psychology, 89,* 335–342.
Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception & Psychophysics, 41,* 191–201.
Posner, M., Snyder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.
Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32,* 65–97.
Pylyshyn, Z., & Burkell, J. (1990, November). *Change: A primitive feature?* Paper presented at the annual meeting of the Psychonomic Society, New Orleans.
Pylyshyn, Z., & Storm, R. (1988). Tracking multiple independent targets: Evidence for both serial and parallel stages. *Spatial Vision, 3,* 179–197.
Remington, R., & Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics, 35,* 393–399.
Sagi, D., & Julesz, B. (1984). Detection vs. discrimination of visual orientation. *Perception, 13,* 619–628.
Saltzman, I., & Garner, W. (1948). Reaction time as a measure of the span of attention. *Journal of Psychology, 25,* 227–241.
Schlosberg, H. (1948). A probability formulation of the Hunter-Sigler effect. *Journal of Experimental Psychology, 38,* 155–167.
Siegler, R. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116,* 250–264.
Simons, D., & Langheinrich, D. (1982). What is magic about the magical number four? *Psychological Research, 4,* 283–294.
Svenson, O., & Sjoberg, K. (1978). Subitizing and counting processes in young children. *Scandinavian Journal of Psychology, 19,* 247–250.
Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 40A,* 201–327.
Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.
Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost filtering. *Perception & Psychophysics, 33,* 527–532.
Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 459–478.
Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.
Trick, L. (1987, June). *Subitizing and canonical pattern.* Paper presented at the annual meeting of the Canadian Psychological Association, Vancouver, British Columbia, Canada.
Ullman, S. (1981). *The interpretation of visual motion.* Cambridge, MA: MIT Press.
Ullman, S. (1984). Visual routines. *Cognition, 18,* 97–159.
Van Oeffelen, M., & Vos, P. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition, 10,* 396–404.
Ward, R., & McClelland, J. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 664–672.
Warren, H. (1897). The reaction time of counting. *Psychological Review, 4,* 569–591.
Wolfe, J., Cave, K., & Franzel, S. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Received January 3, 1991
Revision received June 10, 1992
Accepted June 11, 1992 ■