Vision Res. Vol. 33, No. 18, pp. 2757–2772, 1993
Printed in Great Britain. All rights reserved

0042-6989/93 $6.00 + 0.00

Can We Attend to Large and Small at the Same Time?

BART FARELL,* DENIS G. PELLI*

Received 30 December 1991; in revised form 1 January 1993

Evidence from several sources suggests that visual attention is tuned to stimulus scale. This tuning impairs performance when the observer must attend to more than one scale at a time. In experiments originally designed to measure the bandwidth of attention to stimulus scale, we have found tasks in which observers show no attentional tuning for scale. In separate blocks, observers located or identified targets in arrays of elements, either numbers in arrays of letters or static squares in arrays of flashing squares. Each display contained two arrays, either of the same scale or of different scales. The accuracy of *locating* the target element was lower in mixed-scale than in single-scale displays, but the accuracy of *identifying* the target was unaffected by a mixing of scales within the same display. This holds for both high-level discriminations—numbers vs letters—and low-level discriminations—static vs flashing. Thus, at least for identifying, one can attend to large and small at the same time. The difference in bandwidth between "what" and "where" implies that stimulus identification is not dependent on prior localization.

Attention Channels Size scaling Locating

INTRODUCTION

Suprathreshold vision is remarkably scale invariant. Over a wide range of stimulus sizes, visual performance tends to be constant (e.g. Egeth, 1977; Legge, Pelli, Rubin & Schleske, 1985; McCann, 1978; Parish & Sperling, 1991). This indifference to scale undoubtedly owes much to the existence of multiple spatial frequency channels. How the visual system responds to *changes* of scale, though, is uncertain. Are stimuli of different scales processed independently, or does a change of scale across either space or time require the visual system to make some adjustment in order to effectively process information at both scales? The results of several studies suggest that the bandwidth of visual attention is narrow and that scale invariance is a sequential phenomenon, achieved by shifting the instantaneous scale band to track scale changes, much as light adaptation tracks changes in luminance. However, unlike light adaptation, which is determined by the stimulus, attention to scale, like other aspects of attention, is thought to be largely volitional.

There are two kinds of evidence showing that attention is narrow-band: matching of stimuli of different sizes and searching in mixed-scale displays. In the matching task, observers are asked to ignore size differences. Reaction times for discriminating matching and nonmatching stimulus pairs generally increase linearly with the ratio of linear stimulus sizes (Bundesen & Larsen, 1975; Bundesen, Larsen & Farrell, 1981; Cave & Kosslyn, 1989; Howard & Kerst, 1978; Larsen, 1985; Larsen & Bundesen, 1978; Sekuler & Nash, 1972). Reaction times also increase linearly with the mismatch between expected and presented stimulus sizes (Larsen & Bundesen, 1978). These results have been interpreted in terms of a "zooming" operation analogous to the "mental rotation" of stimuli that have irrelevant orientation differences (Shepard & Metzler, 1971). In "zooming", observers are assumed to process one stimulus size at a time and to select a different size by passing attentionally through intermediate sizes.

The second kind of evidence for scale specificity comes from Sperling and Melchner's (1978) study of visual search in briefly presented arrays of characters. Arrays contained two target digits: a small digit among centrally located small letters, and a large digit among large letters located in surrounding positions. Observers were instructed to distribute their attention in various ratios between the tasks of searching for the small target and searching for the large target. The results showed that for the most part observers failed to comply with the attentional instructions. Instead of devoting 75% of their attention to the small scale and 25% to the large scale, for example, observers appeared to devote all their attention to the small scale on 75% of the trials and to the large scale on the remainder. Sperling and Melchner concluded that the ability to attend to more than one scale at a time is severely limited and that processing stimuli of different scales requires switching attention

---

*Institute for Sensory Research, Syracuse University, Syracuse, NY 13244-5290, U.S.A.

between them. They estimated the time needed for such a switch at about a quarter of a second.

Stimuli at near-threshold contrasts present a different picture. Uncertainty in spatial frequency does reduce the detectability of sinusoidal gratings (Davis & Graham, 1981; Davis, Kramer & Graham, 1983) and lowers near-threshold identification accuracy (Yager, Kramer, Shaw & Graham, 1984), but independent-channels models can account for these effects without assuming a limited attentional capacity (Yager *et al.*, 1984).

That scale is among the stimulus dimensions that can be attentionally selected is not surprising. That attentional selection of a single scale is mandatory, as implied by the suprathreshold studies, has important and unexplored implications for the study of visual performance. These experiments suggested to us that it might be possible to specify attentional limits in terms of properties of spatial frequency channels and to employ the psychophysical techniques developed for studying visual channels to examine effects of attention on perception.

This led us to attempt to measure the scale bandwidth of attention and the relation between attending to stimulus size and attending to spatial area. Our experiments instead led to results that conflict with the original interpretation of those cited above. In assessing the effects of mixing scale, the previous work did not distinguish between judgments of identity and location. We find that the processing of "what" and "where" do not respond in the same way to scale differences. Whereas the bandwidth for "where" judgments is narrow, the bandwidth for "what" judgments is wide: stimuli that differ in scale can be identified, but not located, independently.

Our experiments examined the effects of presenting information at multiple scales. Scale varied across space, not time. The experiments assessed bandwidth limitations by comparing the accuracy of processing information displayed at either of two scales and at both scales at once. Two different tasks were used: identifying or locating the target stimulus. The two tasks were assigned in separate blocks of trials. In all cases the target was embedded within an array of distractors. In all but one experiment (Expt 5), the arrays were related by a scale factor, so that both the size of array elements and their spacing were proportional to the linear size of the array.

METHODS

The effect of presenting multiple scales within the stimulus display was investigated using three experimental methods.

> *Partial report:* four rows of letters appeared briefly, followed by an auditory cue designating the row observers were to report (see Fig. 1).
>
> *Digit search:* a randomly selected digit appeared in an array of letters (see Fig. 2). Observers reported either the digit's identity or its location. In one version of the search display there was one digit target; in another there were two and observers responded to both. Further experiments used variants of this basic search method.
>
> *Flickering checks:* the target was a static square embedded in a randomly flickering checkerboard; observers either located the target or identified it by its color: black or white (see Fig. 3).

Displays appeared on an Apple High-Resolution Monochrome monitor controlled by a Macintosh II computer. The display area measured 21.4 cm horizontally and 16 cm vertically and was refreshed at 67 Hz, non-interlaced. There were 30.3 pixels/cm (77 pixels/in.) vertically and horizontally. In specifying fonts on this display, the Macintosh operating system defines a typographer's *point* as 1 pixel (i.e. 1/77 in.), rather than the standard 1/72.27 in., and we follow this Macintosh convention. Accurate control of stimulus timing and contrast luminance was achieved through use of the publicly available Video Toolbox software (Pelli & Zhang, 1991). In all experiments the stimuli appeared against a gray background luminance of 73 cd/m^2. Details of the displays and procedures are presented below for each experiment separately.

Observers other than the authors were graduate and undergraduate students at Syracuse University. All observers had normal or corrected-to-normal acuity and in each experiment at least one observer was highly experienced in psychophysical tasks and at least one had no prior experience. Observers other than the authors were unaware of the purposes of the experiments. There were three observers in Expts 1 and 5, six in Expt 2, four in Expt 3, and five in Expt 4. Two observers (JW and BF) served in all experiments.

Experiment 1: Partial report

Displays for the partial report experiment are of the type shown in Fig. 1. On each trial four rows of four letters were presented for a duration of 75 msec (five frames). This letter array was centered on the monitor and presented with abrupt onset and offset. To avoid words and pronounceable letters strings and to limit the range of letter heights and widths, the letters A, E, I, O, Q, U, W and Y were excluded. Letters were randomly assigned to the cells of the array with the constraint that no letter appear more than once in any row. There were three scale conditions, which were run in separate blocks of trials: small, large and mixed scale. In the small condition [Fig. 1(a)], the letter size for all rows was 18 point. In the large condition [Fig. 1(b)], displays were simply scaled-up versions of the small displays, but letters were always rendered at the full resolution of the screen. The scale factor was either 2, 3, or 4, applied in separate sessions. Displays in the mixed condition [Fig. 1(c)] contained small letters in the middle two rows and large letters in the outer two.

Observers were required to report the letters of the single target row in order, from left to right. The target row was designated by both auditory and visual cues. Both cues began 225 msec after termination of the display; the tone continued for 120 msec and the visual cue continued until the observer responded. The upper row was cued by the highest-frequency tone (2000 Hz) and those below it by tones of successively lower frequencies: 1000, 400 and 200 Hz. The visual cue consisted of black ellipses abutting the edges of the screen to the left and right of the target row. The presence of the visual cue makes the task seem easier, but had no effect on the results.

The font was upper-case Bookman, a proportionally spaced font with serifs (supplied by Adobe and imaged by the Adobe Type Manager). We overrode the proportional spacing in our displays so that adjacent letters were separated by a constant horizontal center-to-center distance. Viewing distance was either 57 or 114 cm. At 57 cm distance, letters of point size 18 were 0.40° high and a mean of 0.38° wide. Rows of 18-point letters were 1.84° wide, and row spacing (center-to-center) was 0.84°. Corresponding values for larger letter sizes can be derived from the appropriate scale factor. For mixed-scale displays the two inner rows were the same in size and location as the two inner rows in the small condition and the two outer rows were the same in size and location as those in the large condition. A central

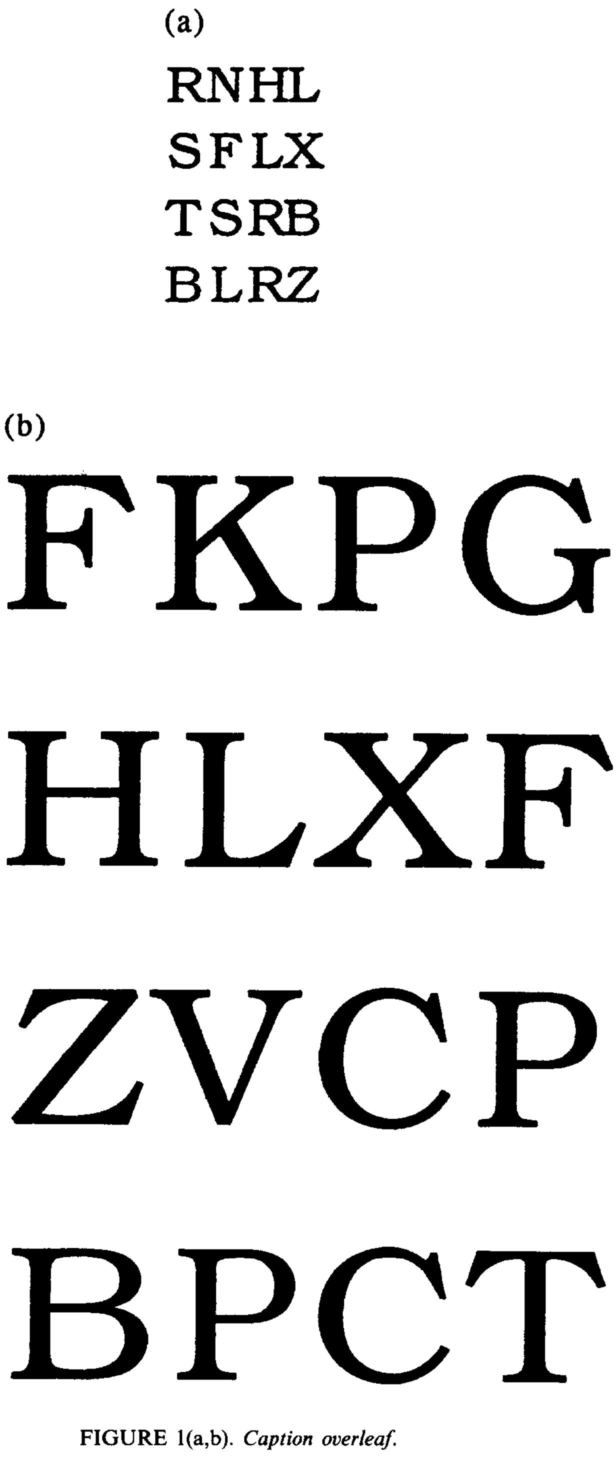

FIGURE 1(a,b). *Caption overleaf.*

FIGURE 1. Displays for partial report. (a) Small condition; display subtense 1.8 × 2.8°. (b) Large condition; display subtense 7.4 × 11.4°. (c) Mixed condition; subtense same as large display.

fixation point preceded display presentation. The luminance of the letters was 10 cd/m^2 and that of the background was 73 cd/m^2.

The row spacings in the displays were set so as to meet the following criterion: when the target row was *precued* with a 120 msec tone beginning 200 msec before display onset, corresponding rows of large and small displays were reported with approximately the same accuracy.

Observers were informed in this experiment and in Expts 2, 3 and 4, that targets could occur at all display positions with equal probability, and they were instructed to monitor all positions without regard to scale.

The three scale conditions were run in separate blocks of 64 trials each. Approximately 12 blocks were completed per daily session, and there was an average of eight sessions per observer. Observers typed their responses and were constrained to respond with no more than four letters, the number of letters in each row of the array. After responding, the observers were shown the target letters on a second monitor.

Experiment 2: Digit search

Observers identified or located a target digit embedded within an array of distractor letters, as illustrated in Fig. 2. The digit was selected randomly from a population of nine digits, "0" being excluded. Neither the digit's identity nor its location was known in advance. The letters excluded from the partial report experiment were excluded here also. Displays typically contained four rows of eight characters each, displayed in the Bookman font with constant spacing. In mixed-scaled arrays, the scale difference occurred between left-hand and right-hand arrays, as illustrated in Fig. 2(c). A replication of the experiment included four blocks of trials: both arrays large, both small, large array on left and small on right, and small array on left and large on right.

There were two methods of presentation. In one, a display was flashed briefly (six frames, 90 msec) on the screen. In the other, a sequence of arrays appeared in rapid succession, each overwriting the one before it, and all but one consisting of letters only, a method first used by Sperling, Budiansky, Spivak and Johnson (1971). Sequences were 12 displays long. The display bearing the target was randomly selected from the middle six of the 12 displays in the sequence (i.e. displays 4, 5, 6, 7, 8, 9). Each display was presented for 90 msec (six frames) and succeeded on the next frame by the following display. The luminance of the letters was 10 cd/m^2 and that of the background was 73 cd/m^2.

For target identification, observers typed a number on the computer's keyboard. For target localization, observers moved a mouse-controlled pointer to one of the possible target positions and clicked a button on the mouse. Each of these positions was marked at its center by a 2 × 2 pixel dot, and the dot closest to the cursor was taken to be the observer's response. Observers received feedback about correct target identity or location on each trial.

Experiment 3: Flickering checks

The display consisted of two laterally adjacent 3 × 3 arrays of squares (checks), as illustrated in Fig. 3. In the multiframe version of the display, each check was initially painted black or white with equal probability and on every subsequent frame (67 Hz rate) each nontarget check had a 50% probability of changing state, black becoming white and vice versa. There was a single static check among the flickering ones and this was the target

FIGURE 2(a,b). *Caption overleaf.*

FIGURE 2. Digit search displays, consisting of two adjacent 4 × 4 arrays of characters. (a) Small condition; display subtense 3.8 × 2.8°. (b) Large condition; display subtense 15.2 × 11.4°. (c) Mixed condition, with display made up of one small and one large array.

that observers either located or identified as black or white. The location task is similar to Verghese and Pelli's (1992) "find the dead firefly" task.

In one condition the two checkerboards were both made up of small checks and in another condition both contained large checks. In a third, mixed, condition, one checkerboard contained small checks and the other contained large checks, as shown in Fig. 3. Each replication of the experiment comprised four blocks of trials: both checkerboards large, both small, large checkerboard on left and small on right, and small checkerboard on left and large on right. The linear size ratio of large and small checks was 6:1. The smallest checks subtended 0.27°, and the largest checks 1.61°. The luminance of the white checks was 110 cd/m^2 and that of the black checks was 10 cd/m^2; the background was 73 cd/m^2.

The flickering checkerboard was presented for 10 frames for one observer and for 12 frames for two others. The first and last frames of the checkerboard were identical. After flickering ceased, the now frozen array of checks remained on the screen as observers responded to the target's location or identity. This they did in the case of location by positioning a mouse-controlled arrow over the check they thought was the target check and then clicking the mouse's button. In the identification task observers responded by positioning the arrow on one of two on-screen buttons labeled "black" and "white". Observers received feedback about the accuracy of each response and, in the location task, were also shown the correct location.

A two-frame checkerboard was also used. The checkerboard was previewed for 0.6 sec and on the next 15-msec frame all checks but the target reversed color between white and black. The checkerboard was removed from the screen on the following frame and replaced after a 0.5 sec delay by dots marking the checks' centers for location responses and by the on-screen buttons for identification responses.

Experiment 4: Dual-digit search

This digit search experiment sought evidence of attentional division or switching between scales by requiring observers to search a display for two targets. This evidence takes the form of conditional probabilities of correctly responding to one target given a correct response to the other target, normalized by the overall proportion correct. One target appeared in each of the two arrays of the display, one in the left-hand array and

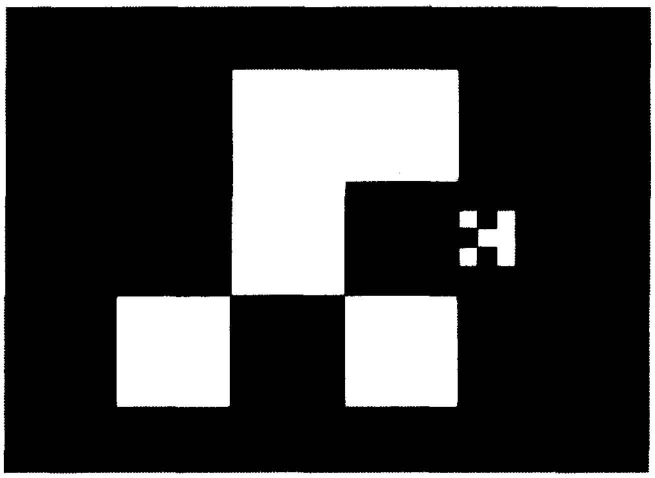

FIGURE 3. Mixed-scale check display, consisting of two adjacent 3 × 3 checkerboard arrays. The small array subtends 0.8° × 0.8° and the large array subtends 4.8° × 4.8°. Single-scale displays were made up of two small arrays or two large arrays.

the other in the right-hand array; in multiframe conditions, the two targets appeared simultaneously. On each trial observers were required to identify the two targets or to locate them. Responses to the left- and right-hand targets could be in either order. In other respects the methods and displays were those of Expt 2.

Experiment 5: Size uncertainty in search

The methods of Expts 1–4 encourage observers to ignore differences of scale. Showing that observers can ignore scale variation would leave unanswered the question of whether they can selectively attend to one scale in multiscale displays. This experiment tests whether scale tuning is possible. For this purpose the task must provide incentive for observers to attend to scale; that is, scale must be relevant, which it was not in the earlier experiments.

The mixed-scale displays in Expts 1–4 cannot distinguish attending to scale (small vs large) from attending to side (left vs right). In order to isolate attention to scale, Expt 5 uses scale differences between members of a single stimulus array rather than between two stimulus arrays. Large and small stimuli with a size ratio of 6:1 were randomly distributed in both time and space within the cells of a sequence of 4 × 4 arrays; again the font was Bookman. Each cell within each array contained with equal probability a large or a small character, as shown in Fig. 4. All arrays had the same area, all cells were the

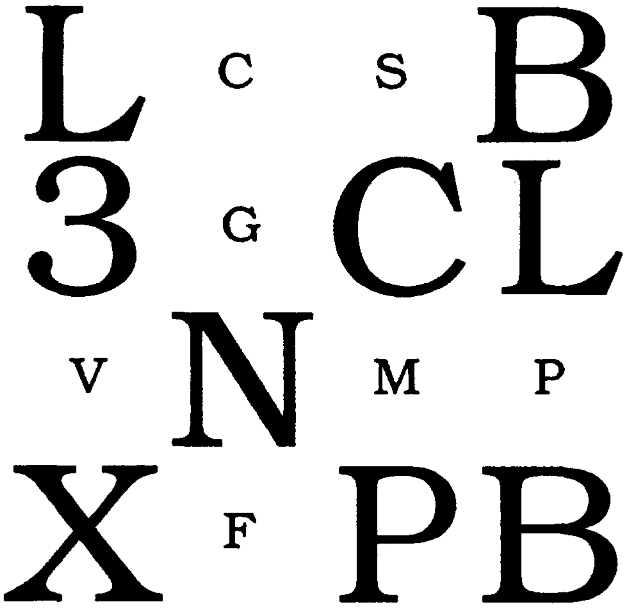

FIGURE 4. Mixed-size display. Display subtense is 8.4 × 8.8° and large-small size ratio of 4:1. Large and small characters had equal probabilities of occurrence within each array and were placed at random within the first arrays in the sequence. Each small character remained in view across two successive arrays and each large character appeared within a single array. The target's probability of being large vs small was varied.

same size, and each character regardless of size was centered within its cell. We call these *mixed-size displays.*

The target was any digit between 1 and 9. There was a single target, a digit, within each sequence of letter arrays. The probability that the target was a particular size, large or small, was varied between blocks of trials. Observers knew this probability in advance. In different blocks of trials the target was small with a probability P of 0, 0.33, 0.5, 0.67 and 1 for two observers and 0, 0.25, 0.5, 0.75 and 1 for a third observer. The target was large with probability $1 - P$. If attending to size benefits performance, one would expect observers to attend to the more probable target size or to attend to a greater extent to this size than to the less likely size.

In mixed-size displays, large letters were closer to fixation, in units of character size, than were small letters. To equate performance for large and small targets in these displays, we varied the exposure duration of the characters of the two sizes. Approximate equality occurred at a 2:1 ratio of durations, 90 msec for large characters and 180 msec for small (6 and 12 frames, respectively). Since the large and small characters were spawned with equal probability, but the small ones remained on screen twice as long, the sequence of 12 arrays in each trial spawned twice as many large characters as small characters.

RESULTS

For Expts 1–4 a comparison was made between accuracies in single-scale and mixed-scale displays. The outcomes of these comparisons depended on the task: identifying or locating. Jumping ahead for a moment, Fig. 11 summarizes these results, plotting the ratio of percent correct on mixed- and single-scale displays for each stimulus and task. In all of the experiments this ratio is about 1.0 for the identifying; that is, the accuracy of identifying is the same in mixed- and single-scale displays. For locating the ratio is about 0.75: the accuracy of locating is 25% worse in mixed-scale than in single-scale displays.

Graphs that appear below for the first four experiments show the probability of a correct response to a single-scale display plotted against the probability of a correct response to a mixed-scale display. Data points for identifying cluster around the diagonal, indicating similar accuracy in mixed- and single-scale displays. Data points for localizing fall below the diagonal, indicating superior accuracy when there is only a single scale. Data for large and small targets are plotted separately, to reveal any difference in the effects of scale mixing due to unequal focusing of attention between the two scales.

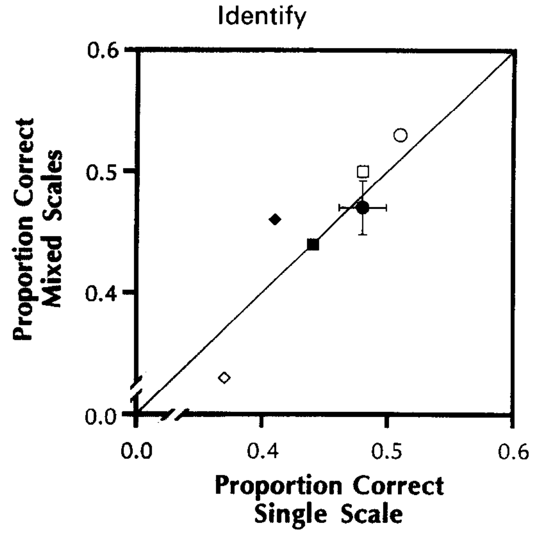

FIGURE 5. Partial report data (Expt 1). The proportion of correct identifications of cued rows in single-scale displays (x-axis) is plotted against that in mixed-scale displays (y-axis). Data for each of the three observers are plotted with a different symbol. Open symbols are for small letters; solid symbols are for large letters. Error bars indicate ± 1 typical standard error. The size ratio was 4:1. Responses were scored without regard to order of report. The diagonal line represents equal performance on single- and mixed-scale displays. Circles are for observer PV, squares for BF and diamonds for JW.

Partial report

Partial report is an identification task, and it showed no effect of mixing scales. Probabilities of correct responses with a 4:1 size ratio appear in Fig. 5. Letters were scored without regard to the order in which they were reported. Accuracy with single-scale displays is indicated on the horizontal axis and accuracy with mixed-scale displays is given on the vertical axis.* Error bars show ±1 typical standard error; as in all the experiments, these were computed across the eight blocks of trials per session and averaged across sessions. The diagonal line represents equal performance with single- and mixed-scale displays. None of the data points differ significantly in either the horizontal or the vertical direction from the diagonal line of Fig. 5. Each observer's overall probability correct for single-scale displays was within 1% of that for mixed-scale displays. A similar graph resulted when responses were scored as correct only if the target letters were reported in the correct order. The data are consistent with the null hypothesis; performance does not differ between single- and mixed-scale displays.

The lack of a significant effect of mixing scales argues against the notion that observers can attend to only one of these scales at a time. The scale of the targets was unknown before the cue. Switching attention to the cued scale would therefore be expected on at least half of the trials (half if attention were focused to one of the two scales before the arrival of the cue, and more than half if attention were in a non-selective state or focused on an intermediate scale). The time taken to switch attention should mimic a further delay of the cue: both events would reduce performance by delaying

---

*The four rows of the arrays were cued equally often. However, only half the data from single-scale displays were of interest: data from the outer two rows of large arrays and from the inner two rows of small arrays. Only these rows are the same in single-scale and mixed-scale arrays.

processing from a decaying memory trace (Sperling, 1960).

Digit search

When observers were required to identify, the digit-search experiment showed no overall effect of mixing scales. When observers were required to locate, performance suffered when scales were mixed. Note that in this and subsequent experiments, the relative level of identification and localization accuracies are not relevant to the effect of mixing scales; these are separate tasks, though they use the same stimuli.

Target identification accuracies are shown in Fig. 6(a). The scale ratio was 4:1, and data were collected using sequential presentations of 12 displays per trial. Each of the six observers contributed two data points: open symbols for the small scale and solid symbols for the large scale. Data for the two scales are similar and only one of the data points (open triangle, observer CS) deviates significantly from the diagonal line marking equivalent performance on single-scale and mixed-scale displays. The decline in identification accuracy shown by observer CS for small targets in mixed-scale displays was not accompanied by a significantly enhanced accuracy for large targets. Thus scale mixing caused a general decline in accuracy for this observer, rather than a trade-off between the large and small scales.

Localization data appear in Fig. 6(b). The data points for five of the six observers are below the diagonal line, indicating that targets were located less accurately in mixed-scale than in single-scale displays. For these observers, the probability of locating a target in single-scale displays averaged 30% greater than in that in mixed-scale displays. All of the data points differ significantly from the diagonal line except for two: large targets for observer JW (solid diamond) and small targets for observer JG (large open circle), whereas the combined data for large and small targets differ significantly from the diagonal for all five observers. Observer CS (triangles) combined superior localization of large targets in mixed-scale displays with near-chance localization of small targets in these same displays. This observer located large and small targets with comparable accuracy in single-scale displays. Her data could result from attending exclusively to the large array when scales are mixed; she reported using this strategy because of an inability to monitor both scales at once while locating.

An analysis of location errors of three observers showed that the average size of offset between the target's location and the observer's response was greater in mixed-scale displays than in single-scale displays. This is also true of the offsets of responses that were incorrect but nevertheless were on the same array as the target. Moreover, observers more often missed the target's array when scaling was mixed, i.e. responded on the right side of the display to a target on the left, or vice versa. Therefore observers were more often ignorant of target location in mixed-scale displays than in single-scale displays, resulting in more guesses, rather than merely making near-miss motor errors to accurately perceived target locations.*

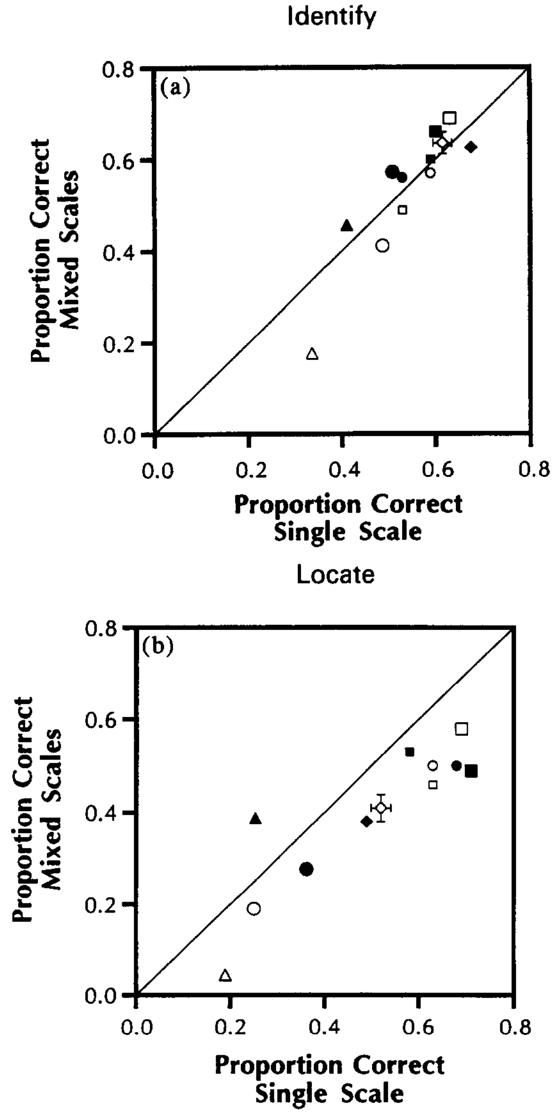

FIGURE 6. Digit search data (Expt 2). (a) Proportion correct identifications. (b) Proportion correct localizations. Data are from the sequential-presentation method and are plotted as in Fig. 5. The scale ratio was 4:1. Large circles are for observer JG, large squares for BF, diamonds for JW, triangles for CS, small circles for JS, and small squares for SS.

The target and distractor elements in this and the other experiments reported here were arranged in regular arrays and matching grids of locations were used to mediate localization responses. In an experiment similar to the present one, these features were eliminated: characters were positioned quasi-randomly at low density

---

*An observer ignorant of the targets location in a mixed-scale display can still use knowledge of the target's size to confine guesses to the correct half of the display. This knowledge of the target's size should double a rational observer's probability of correctly guessing its location, relative to the probability of a correct guess in single-size displays. No counterpart of this probability effect exists in the identification task. This suggests that in this experiment (and in the following one as well) we might have underestimated the degree to which mixing sizes impairs localization accuracy, which strengthens our conclusion that mixing scales impairs locating.

(four per display) and localizing involved placing an outline of the target at the remembered location. Again, localizing, but not identifying, suffered as a result of mixing scales.

A comparison of single displays, a sequence of displays, and various presentation durations (90, 75 and 60 msec) showed that these variables affected overall performance but did not influence the relative accuracy for single- and mixed-scale displays in either the identifying or locating task.

Flickering checks

The task of identifying the target check as black or white did not show an overall effect of mixing scales. Figure 7(a) shows the proportion correct for the identifying task. There was a 6:1 linear ratio of large to small. Error bars show ±1 standard error. As in the previous graph, the diagonal line indicates equal performance in single- and mixed-scale displays. One observer (JW) was poorer at identifying small checks when scales were mixed (open diamond).

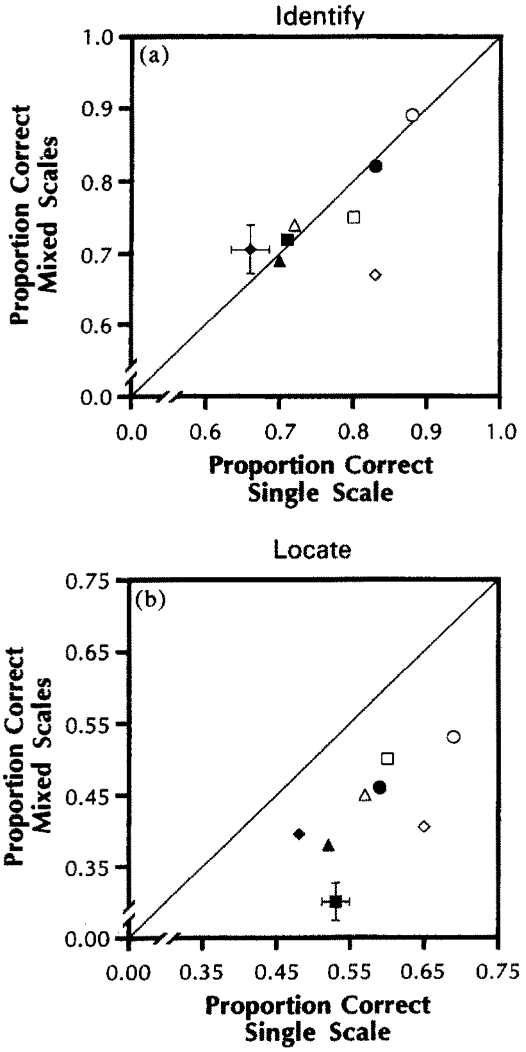

FIGURE 7. Flickering checks data (Expt 3). (a) Proportion correct identifications. (b) Proportion correct localizations. Data are from the two-frame method and are plotted as in Fig. 5. The size ratio was 6:1. Circles are for observer CB, squares for BF, diamonds for JW, and triangles for DZ.

When the task was to locate the target check, mixing scales did have a significant effect for all observers and for both large and small targets. These data appear in Fig. 7(b). Targets were more accurately located in single- than in mixed-scale checkerboards.

Dual-digit search

In this search experiment there were two targets, one in each of the two arrays of a display, and the observer responded to both of them. When identifying single targets, an attentionally-limited observer might conceivably show little or no effect of mixing scales. If constrained to fully attend to only one scale at a time, the observer would attend to a smaller number of array elements in mixed-scale displays than in single-scale displays. The advantage of attending to fewer elements could in principle offset the cost of not fully attending to the other scale. Because of this possibility, the dual-digit search task is a critical test of the limited-bandwidth hypothesis for identification.

There are two ways in which a limited attentional bandwidth for stimulus scale can reduce performance when targets differ in scale. In one way, attention is focused on the scale of one target and must change its focus before information at another scale can be processed. In a display that allows only limited processing time this will result in a tradeoff between the accuracies for target of different scales. Thus the conditional probability (of a correct response to one target, given a correct response) to the other target, will be a smaller proportion of the unconditional probability (of a correct response) when scales are mixed than when they are the same. This proportion, which we will call the *conditional proportion* of correct responses, has an expected value of 1.0 when responses to the two targets are independent. It will exceed 1.0 when the two targets are processed synergistically and will be less than 1.0 when the processes compete, e.g. for the limited-bandwidth of attention.

The second way that a limited bandwidth can reduce performance is if attention is not focused on either scale of the display, but rather focused on an intermediate scale. The conditional proportion in this case need not differ between single-scale and mixed-scale displays, but accuracy will be lower in mixed-scale displays because attention is not focused optimally for either target.

Figure 8(a) shows that the probability of correctly identifying a target in mixed-scale displays is very nearly the same as in single-scale displays despite a 4:1 size ratio. None of the distances between the points on the graph and the diagonal line is significant. The conditional proportions for correct identifications appear as solid symbols in Fig. 9(a). Because a target might be correctly identified by guessing at random, the conditional proportions would be above 0.0 even for an observer who could attend to only one of the scales of the display and so had to guess the identity of the target at the other, unattended, scale. If our observers had guessed in this way, their maximum expected mixed-scale conditional proportions are those plotted by open

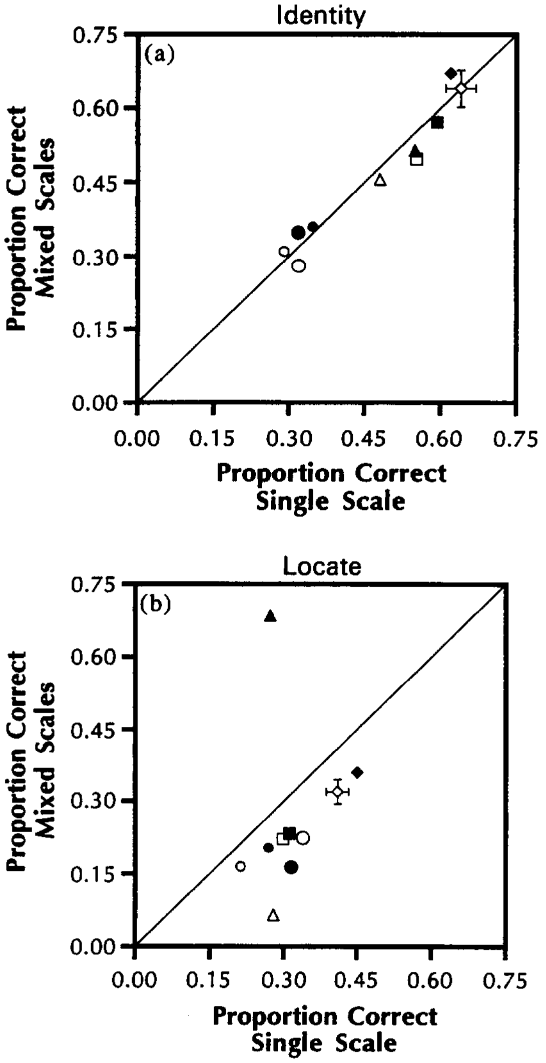

FIGURE 8. Dual-digit search data (Expt 4). (a) Proportion correct identifications. (b) Proportion correct localizations. Data are from the sequential-presentation method and are plotted as in Fig. 5. The scale ratio is 4:1. Large circles are for observer JC, squares for BF, diamonds for JW, triangles for CS, and small circles for NL.

symbols in Fig. 9(a). The observed and expectedly y-axis values in this figure differ markedly, indicating that attending to one scale does not block attention to the other scale.

The solid symbols in Fig. 9(a) are also all < 1.0, indicating competition (i.e. trade-off) between identifying the target in the left-side array and identifying the target in the right-side array. However, they are close to the diagonal line and so give no indication of a systematically larger trade-off when the two arrays have different scales rather than the same scale. The limited bandwidth hypothesis predicts that one or both of the statistics shown in Figs 8(a) and 9(a) will be higher for single-scale displays than for mixed-scale display, and this is not the case. The obtained results are expected if identifying is insensitive to scale differences.

Figure 8(b) shows results for locating. The overall probability of correctly locating was lower in mixed-scale than in single-scale displays, consistent with earlier results. All the data points but 1 (observer NL, small open circle) are significantly distant from the diagonal. The conditional proportions for locating appear in Fig. 9(b) as solid symbols. For several observers the conditional proportions approached 1.0, indicating little trade-off between localizing the two targets. Values for other observers are considerably below 1.0, indicating a trade-off. However, for three of the five observers, the extent of the trade-off is independent of any scale difference between the targets; their data points (plotted as large circle, square and diamond) fall near the diagonal line. Finally, in all cases but one, the conditional proportions are well above the values expected if observers could attend to only one scale per display, shown as open symbols. The exception, as expected, is CS, shown by open triangles, for whom locating at one scale is incompatible with locating at a different scale.

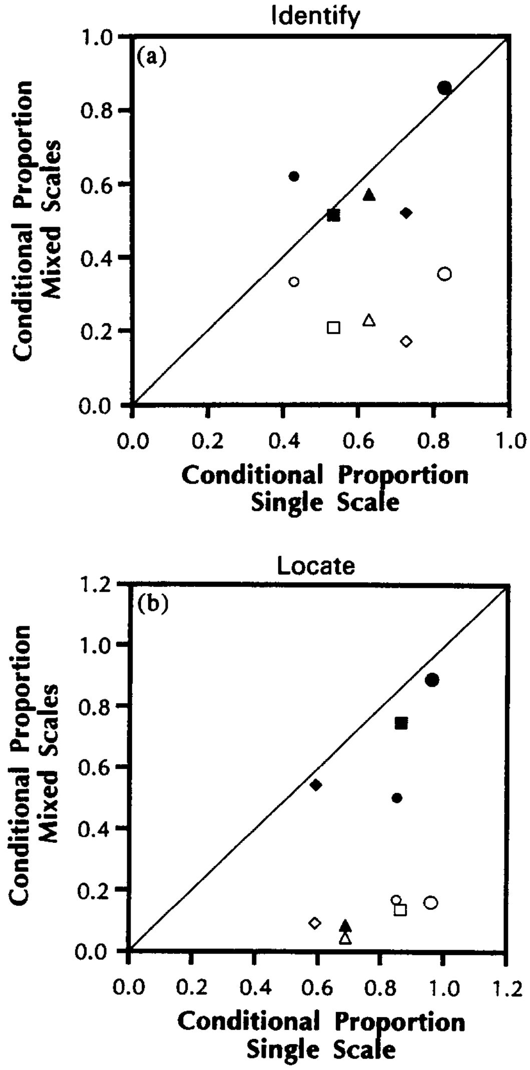

FIGURE 9. Conditional proportions for dual-digit search (Expt 4). (a) Conditional proportion of correct identifications. (b) Conditional proportion of correct localizations. There is a different symbol for each of the five observers. Solid symbols plot the probability that the observer responded correctly to one target given that he or she responded correctly to the other target, normalized by the unconditional probability of a correct response. Open symbols show the maximum conditional proportions expected if correctly responding to a target of one size reduced the observer to responding at random to the target of the other size. The assignment of symbols to observers is the same as in Fig. 8.

The data indicate that most of our observers attended to an intermediate scale when locating within mixed-scale displays. This intermediate scale tuning is optimal for neither of the display scales. For this reason mixed-scale displays make target localization more difficult than do single-scale displays, even though the two display types show similar trade-offs between the two targets.

The identification task shows no evidence that observers are limited to attending to only one of the array scales or that they attend to an intermediate scale. The results reject the hypothesis of a bandwidth limitation in identifying.

Size uncertainty in search

The proportion of targets correctly identified in Expt 5 is shown in Fig. 10(a). The data are plotted as a function of the probability of the target's size. This probability, which for different observers ranged from 0.25 or 0.33 to 1.0, generally had a statistically insignificant effect on identification accuracy.

The results of the previous experiments showed that observers can ignore scale when identifying, not that they must do so. If attending to a particular size aids identification at that size, accuracy in this experiment should increase with probability or at least be greater when uncertainty about target size is absent ($P = 1.0$) than when it is greatest ($P = 0.5$). The data indicate that the benefits of attentional size tuning for identification accuracy must be quite small.*

Localization data were also collected and appear in Fig. 10(b). Note that they show no more evidence for attentional selection of size than do the identification data. This indicates that the scale specificity of localizing seen in Exps 2–4 is tied to the *spacing* of the array elements—i.e. tied to the scaling of the positions of the elements—not to their sizes.

DISCUSSION

The results show that attention can span scale differences across space. From earlier research we expected to find a limited attentional bandwidth for scale differences regardless of the nature of the experimental task. This bottleneck would affect all tasks similarly if it were a limitation imposed by a common visual pathway. Instead we find that scale tuning is task-dependent rather than being a general attribute of visual attention. Determining *what* an element is is wide band. However, determining *where* an element is is narrow band and scale specific. These results imply that discriminating targets and distractors must, like identifying, be wide band, as discussed below.

Across our experiments the accuracy of locating a target in mixed-scale displays averaged 25% below that in single-scale displays, whereas the accuracy of identifying a target in the same displays showed no effect of scale mixing. Figure 11 summarizes the results of the experiments, showing the ratios of mixed-scale to single-scale accuracy for identifying and locating. In a companion paper, Verghese and Pelli (1993) show the scale-tuning function of the locating response and present a simple model to quantitatively account for the results.

Sperling and Melchner (1978) found that observers attended to only one scale when searching mixed-scale displays. The data were the proportion of targets identified correctly. However, observers had to report not only the targets' identities, but also their whereabouts; and they had to rate their confidence in two identification responses as well. Localizations were used to screen for randomly guessed target identifications: a correct identification was scored as correct only if the target was correctly located or mislocated by only one row or column of the array. So Sperling and Melchner's

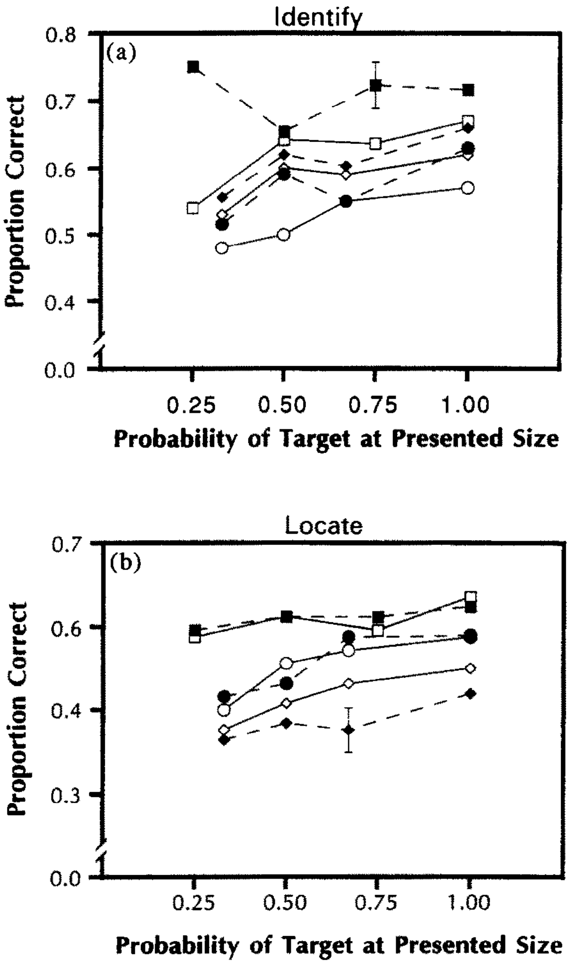

FIGURE 10. Mixed-size digit search data (Expt 5). (a) Proportion correct identifications. (b) Proportion correct localizations. Data are plotted as a function of the probability of target at the presented size. Each of the three observers has been given a different symbol; open symbols are for the small size and solid symbols are for the large size. Size ratio was 4:1. Circles are for observer LW, squares for BF, and diamonds for JW.

---

*Of the observed (statistically insignificant) effect of probability, only part can be attributed to attention. This is because the ideal strategy for the observer is Bayesian, using the prior knowledge of the probabilities of the two target sizes to help choose among candidate targets of different sizes. Thus, even without selectivity attending to size, the observer might show an effect of probability.

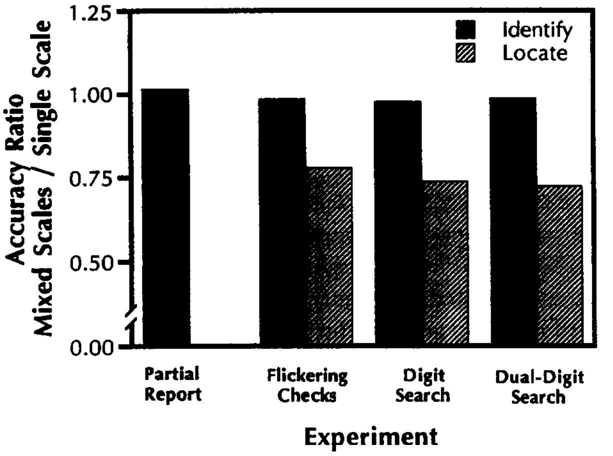

FIGURE 11. Ratios of mixed- and single-scale accuracies for identifying and for locating in Expts 1–4, averaged across observers. Solid bars are for identifying and shaded bars are for locating. Data from observer CS have been excluded because her strategy differed markedly from that of the rest of the observers in the digit search experiment.

observers were processing the location of targets in mixed-scale displays. That their identification data, conditionalized on localization responses, should display effects of scale differences is therefore expected and compatible with our localization data.

We have tested this reasoning in an identify-and-locate task using the flickering checks stimuli. Two observers first located and then identified the target on each trial; they felt the reverse order of responses to be less congenial. The targets were both located and identified more accurately in single-scale than in mixed-scale checkerboards, as expected from Sperling and Melchner's results. Correctly located targets were almost always correctly identified, as Sperling and Melchner also observed.*

Identifying a stimulus and locating it require, at least in part, different kinds of information and this may have both anatomical and behavioral consequences. From neurological evidence Ungerleider and Mishkin (1982) have argued for an anatomical separation between "object vision" and "spatial vision", suggesting that a striate-parietal pathway carries information about location and that a striate-temporal pathway is involved in the processing of object identity. A distinction between object and location processing has also appeared in behavioral experiments of Treisman and Gelade (1980) and Julesz (1981), among others. Our data too imply that an object's identity and its location are at least in part processed separately.

Engel (1974) measured the largest eccentricity at which a brief incremental spot could be detected on a uniform field. He found that there was no difference in the detectability of dots presented at known locations in the visual field and those presented at unknown locations. Similar null effects have been reported by Mowrer (1941), Mertens (1956), and Grindley and Townsend (1968), among others. In addition, Pelli (1981a, b) showed that detectability of a brief line was little affected by a ten-thousand-fold uncertainty about location and time of presentation. These studies show that detection of a brief spot or line is insensitive to positional uncertainty.

Our data show that identifying is insensitive to uncertainty in scale, but that, when locating, people do so with respect to a particular spatial scale. It seems that knowing the identity of an element does not entail a knowledge of its location, whereas knowing the location of an element implies knowing, or at least being influenced by, the scale of the element's possible positions.

The data also imply that discriminating targets and distractors is, like identifying, not scale specific. Both identifying and locating require the discrimination of targets from distractors. Discriminating static and flickering checks or digits and letters must not impose any scale specificity, as otherwise both identifying and locating would show it.†

Is what vs where the right distinction?

We have presented the *what* vs *where* distinction as crucial for the scale specificity of attention. Another possibility is that our identifying and locating tasks have as their crucial difference only the quantitative one of task difficulty. A difficult task might use up resources needed to process information at multiple scales. It would then appear scale specific whereas an easier version of the same task might not. On this hypothesis a sufficiently easy locating task would produce no effect of scale variation and a difficult identification task would show the effect.

To test this possibility we varied task difficulty by changing the number of cells in our flickering checkerboards. With checkerboards whose number of cells differ by a factor of two, we can compare performance for single- and mixed-scale checkerboards that have the same number of cells of a given size. These checkerboards should be similarly difficult at that scale. We went beyond this, extending the number of cells over a

---

*One possibility is that observers combine the identifying and locating tasks by giving precedence to locating; they locate the target and then identify the stimulus at this location. This is a good strategy if identifying by itself provides no information about location. When reporting both target identity and location in this experiment, the tight empirical association between knowing where the target is and knowing what it is has a strong subjective counterpart. Subjectively, identity and location usually seem inseparably bound. In this regard the identify-and-locate task differs from the identify-only task. In the identify-only task there is often the realization that one has little or no awareness of the location of the target (nor the target's size, when sizes are mixed). Requiring observers to report a conjunction of properties, such as identity and location, could have additional consequences. As Treisman and Gelade (1980) have shown, processing conjunctions of two or more stimulus properties can severely restrict attention. Thus observers such as Sperling and Melchner's, who both identify and locate the two targets presented on each trial, might encounter attentional limitations that encourage a trade-off between the two targets, even if both were the same size.

†Whether one must identify a digit in order to discriminate it from letters has been a controversial question. For evidence favoring the view that the discrimination can be "categorical", i.e. not contingent on identifying (see, e.g. Gleitman & Jonides, 1976; Jonides & Gleitman, 1976).

four-fold range, using 2 × 2, 3 × 3 and 4 × 4 checkerboards. The results do not support the task difficulty hypothesis. Increasing the number of cells by a factor of four reduced proportion of correct localizations by a factor of 2.4 and 1.9 for the two observers, but, as shown in Fig. 12, this did not raise the relative difficulty of locating in mixed-scale displays. If anything, the ratio of correct localizations in mixed-scale displays to those in single-scale displays increases for one observer as more checks are added. Identification too was affected by number of cells, though not as much as was locating, and—again contrary to the hypothesis—mixed-scale displays did not become more difficult relative to single-scale displays as more cells were presented. Thus scale specificity seems governed by the type of task—identifying or locating—and not by quantitative variation within or between tasks.

Sagi and Julesz (1985) have suggested that locating occurs in parallel across several display items whereas identifying the same items is a serial process involving shifts of attention. These data concur with ours in supporting the separability of locating and identifying. Clearly, our data would not be in agreement with a sequential-dependency interpretation of Sagi and Julesz's results, whereby a preattentive locating process precedes and spatially guides an attentive identification of the stimuli. However, nothing in their data demands such a dependency and, as suggested below with reference to identify-and-locate tasks, the relation between locating and identifying may well be flexible and under strategic control.

Johnston and Pashler (1990) have recently disputed the claim of earlier studies that identifying and locating are independent processes. In their experiments, they found that correctly identifying a stimulus was strongly linked to correctly locating it. However, this only shows that the two measures can be strongly linked, not that they must be. Treisman and Gelade (1980) found a high probability of locating a target stimulus, given that it had been correctly identified, only for those stimuli that did not pop out from a background of non-targets. These targets required a search. Targets that did pop out were rather poorly located. The linkage found by Johnston and Pashler might also have resulted from the strategy adopted by their observers in order to both identify and locate the target stimulus, as required by the experiment. As discussed above, identification accuracy in identify-and-locate tasks differs from that in identify-only tasks.

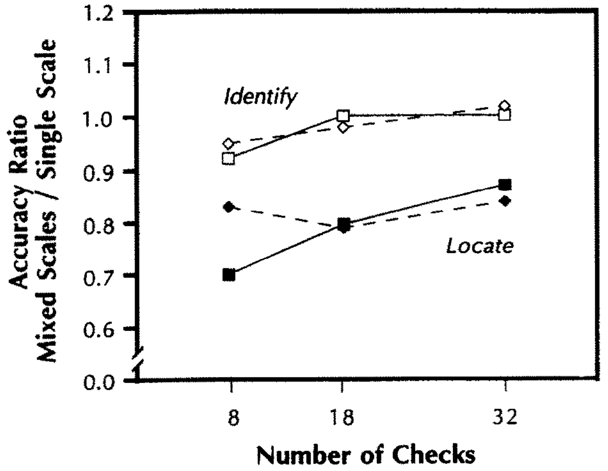

FIGURE 12. Ratio of mixed- and single-scale accuracies for identifying and for locating the static check among background flickering checks, as a function of the total number of checks. Data for each observer are plotted by a different symbol. The open symbols denote performance in the identification task and the solid symbols denote performance in the localization task. The squares are for observer BF and the diamonds for JW. The task-difficulty hypothesis predicts that the ratios will decrease as the number of checks increases, and is rejected by these results.

As conceptually clear as the *what* vs *where* distinction may appear, the two cannot be fully dissociated. For example, identifying stimuli that are defined by different combinations of the same set of components might require that the components be located (Julesz, 1981; Treisman & Gelade, 1980; cf. Farell, 1984). Indeed, the recognition of complex visual displays may depend more on the phase spectrum than on the amplitude spectrum (Farell & Julesz, 1989; Oppenheim & Lim, 1981). Thus *what* and *where* are not always entirely independent stimulus properties, leaving open the question of precisely what difference between the tasks accounts for the difference in scale specificity.

We are currently investing the possibility that location responses depend on a spacially bandlimited visual memory (Pelli & Farell, 1992), whereas over-learned identifications of digits or colors might not require visual memory at all.

Stimulus matching

Stimulus matching experiments, as discussed in the Introduction, have shown effects on reaction time of an irrelevant size mismatch, in support of the notion of a narrow-band attentional tuning to scale. In light of the present data, assessment of these studies clearly hinges on whether they probed identity or location. None of them involved an explicit discrimination of stimulus location. Perhaps, then, the results of matching studies are further instances of the task dependence found in our research. Yet several of the matching studies required discrimination of spatial properties and were not unambiguously concerned with stimulus identity more than with spatial location. In one example, complex patterns that were identical except for irrelevant orientation and size differences were discriminated from those that also differed by a mirror-image reflection (Larsen, 1985). Even in experiments where the mismatching stimuli differed in shape, the complexity and unfamiliarity of the stimuli used might have made the discrimination depend on a test of spatial congruence. Tests of spatial congruence, which seem necessary for discriminating mirror images (Corballis & Beale, 1976; Shepard & Metzler, 1971), might well require "zooming" of size-differing stimuli, thereby accounting for the effect on reaction time. On the other hand, stimulus-matching reaction times are insensitive to size differences when the set of stimuli contains few members and so becomes familiar to observers (Kubovy & Podgorny, 1981; Larsen, 1985). If these familiar stimuli were recognized by absolute identification—i.e. without reference to a comparison stimulus—then congruence testing, and therefore zooming too, would be unnecessary.

Matching experiments, then, bear an uncertain relation to the present study. A direct comparison of stimulus identity matches and stimulus location matches would be needed to bridge the gap.

Temporal resolution and spatial uncertainty

Our stimuli were brief. Differences in temporal resolution between channels responding to large and small scales could have influenced performance. These temporal differences might even have caused observers to process the elements in different ways depending on the scale of the array. It is therefore possible that the nature of the target-distractor discrimination changed between different array scales, so that mixing scales would present the observer with a mixture of kinds of discrimination. This issue of temporal resolution has a counterpart in the spatial domain. However, such caveats lose considerable force in the present experiments because the identifying task, which showed no effect of mixing scales, acts as a control for the locating task, which did show such effects.

The wide bandwidth we attribute to identification may result from parallel processing across many channels, each specific to a different stimulus scale. Consideration of the multiplicity of such channels might in turn suggest that effects of mixing scale could be explained as effects of uncertainty. Indeed, our experiments resemble uncertainty experiments, and identifying and locating might differ in their susceptibility to uncertainty about scale. However, mixing scales can have an effect even when it adds no uncertainty. This is the case in our dual-digit search experiment and in Sperling and Melchner's (1978) study, where targets appeared at both scales on every trial.

CONCLUSIONS

The visual stimulus limits what can be seen but it does not determine what is seen. Observers typically pay attention to some aspects of the visual scene at the expense of others. Previous experiments suggested that mixed-scale displays could reveal the scale-specific nature of visual attention. Initially this project sought to link the existing evidence for scale tuning in attention with spatial frequency tuning in detection. Instead we discovered that scale tuning is task specific. Observers are poorer at localizing (saying *where*) in mixed-scale displays than in single-scale displays, but their ability to identify (say *what*) is unimpaired. This holds for both high-level discriminations—letters vs digits—and low-level discriminations—static vs flashing. Averaged across experiments and observers, locating in mixed-scale displays was only 75% as accurate as in single-scale displays, while identifying was unaffected. At least when identifying, one can attend to multiple scales at the same time. The scale specificity of locating implies that one can identify a stimulus without having explicitly located it. This scale specificity may reflect the narrow bandwidth of visual memory (Pelli & Farell, 1992).

REFERENCES

Bundesen, C. & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 214–220.

Bundesen, C., Larsen, A. & Farrell, J. E. (1981). Mental transformations of size and orientation. In Long, J. & Baddeley, A. (Eds), *Attention and performance IX.* Hillsdale, N.J.: Erlbaum.

Cave, K. R. & Kosslyn, S. M. (1989). Varieties of size-specific visual selection. *Journal of Experimental Psychology: General, 118,* 148–164.

Corballis, M. C. & Beale, I. L. (1976). *The psychology of left and right.* Hillsdale, N.J.: Erlbaum.

Davis, E. T. & Graham, N. (1981). Spatial uncertainty effects in the detection of sinusoidal gratings. *Vision Research, 21,* 705–712.

Davis, E. T., Kramer, P. & Graham, N. (1983). Uncertainty about the spatial frequency, spatial position, or contrast of visual patterns. *Perception & Psychophysics, 33,* 20–28.

Egeth, H. (1977). Attention and preattention. In Bower, G. H. (Ed.), *The psychology of learning and memory* (Vol. 2). New York: Academic Press.

Engel, F. L. (1974). Visual conspicuity and selective background interference in eccentric vision. *Vision Research, 14,* 459–471.

Farell, B. (1984). Attention in the processing of complex visual displays: Detecting features and their combinations. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 40–64.

Farell, B. & Julesz, B. (1989). Globally perceived directional flow in static images. *Perception, 18,* 155–172.

Gleitman, H. & Jonides, J. (1976). The cost of categorization in visual search: Incomplete processing of targets and field items. *Perception & Psychophysics, 20,* 281–288.

Grindley, C. G. & Townsend, V. (1968). Voluntary attention in peripheral vision and its effect on acuity and differential thresholds. *Quarterly Journal of Experimental Psychology, 20,* 11–19.

Howard, J. H. Jr & Kerst, S. M. (1978). Directional effects of size change on the comparison of visual shapes. *American Journal of Psychology, 91,* 491–499.

Johnston, J. C. & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 843–856.

Jonides, J. & Gleitman, H. (1976). The benefit of categorization in visual search: Target location without identification. *Perception & Psychophysics, 20,* 289–298.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290,* 91–97.

Kubovy, M. & Podgorny, P. (1981). Does pattern matching require the normalization of size and orientation? *Perception & Psychophysics, 30,* 24–28.

Larsen, A. (1985). Pattern matching: Effects of size ratio, angular difference in orientation, and familiarity. *Perception & Psychophysics, 38,* 63–68.

Larsen, A. & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 1–20.

Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Research, 25,* 239–252.

McCann, J. J. (1978). Visibility of gradients and low spatial frequency sinusoids: Evidence for a distance constancy mechanism. *Photographic Science and Engineering, 22,* 64–68.

Mertens, J. J. (1956). Influence of knowledge of target location upon the probability of observation peripherally observable test flashes. *Journal of the Optical Society of America, 46,* 1069–1070.

Mowrer, O. H. (1941). Preparatory set (expectancy)—further evidence for its "central" locus. *Journal of Experimental Psychology, 28,* 116–133.

Oppenheim, V. A. & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE, 69,* 529–541.

Parish, D. H. & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise and the efficiency of letter discrimination. *Vision Research, 31,* 1399–1415.

Pelli, D. G. (1981a). The effect of uncertainty: Detecting a signal at one of ten-thousand possible times and places. *Investigative Ophthalmology and Visual Science (Suppl.), 20,* 178.

Pelli, D. G. (1981b). Effects of visual noise. Ph.D. thesis, Cambridge University, Cambridge.

Pelli, D. G. & Farell, B. (1992). Visual memory. *Investigative Ophthalmology and Visual Science (Suppl.), 33,* 1263.

Pelli, D. G. & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31,* 1337–1350. The VideoToolbox software is available by anonymous ftp of Info-mac/Development/SourceCode/video-toolbox from summex-aim.stanford.edu [36.44.06].

Sagi, D. & Julesz, B. (1985). "What" and "where" in vision. *Science, 228,* 1217–1219.

Sekuler, R. & Nash, D. (1972). Speed of size scaling in human vision. *Psychonomic Science, 27,* 93–94.

Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (11).

Sperling, G. & Melchner, M. J. (1978). The attention operation characteristic: Examples from visual search. *Science, 202,* 315–318.

Sperling, G., Budiansky, J., Spivak, J. G. & Johnson, M. C. (1971). Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science, 174,* 307–311.

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

Ungerleider, L. G. & Mishkin, M. (1982). Two cortical visual systems. In Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W. (Eds), *Analysis of visual behavior.* Cambridge, Mass.: MIT Press.

Verghese, P. & Pelli, D. G. (1992). The information capacity of visual attention. *Vision Research, 32,* 983–995.

Verghese, P. & Pelli, D. G. (1993). Scale bandwidth of visual search. *Vision Research.* In press.

Yager, D., Kramer, P., Shaw, M. & Graham, N. (1984). Detection and identification of spatial frequency: Models and data. *Vision Research, 24,* 1021–1035.

Acknowledgements—This work was supported by National Eye Institute grants EY06270 to Bart Farell and EY04432 to Denis Pelli. Preeti Verghese contributed to discussions that provided the original motivation for this work, and provided helpful comments on an earlier draft of this manuscript. We thank George Sperling for discussing several of these ideas with us, and for suggesting the devil's advocate task-difficulty hypothesis that we present in the Discussion. Portions of this study were presented at the 1991 meeting of the Association for Research in Vision and Ophthalmology.

Section 3

Computational Vision