Journal of Experimental Psychology:
Human Perception and Performance
2002, Vol. 28, No. 2, 315–334

0096-1523/02/$5.00 DOI: 10.1037//0096-1523.28.2.315

Temporal Integration Between Visual Images and Visual Percepts

James R. Brockmole, Ranxiao Frances Wang, and David E. Irwin
University of Illinois at Urbana–Champaign

Using a temporal integration task, subjects in 5 experiments were expected to combine information from temporally separated visual presentations. Evidence from these experiments indicated that perceptual information can be integrated with previously generated and currently maintained visual images to form a representation that contains information from each source. Properties and limitations of this integration process were also explored, including the time required to generated the image, the speed at which percepts were integrated with images, and the capacity of the representation. Implications for theories of visual processing and memory are discussed.

Phenomenologically, visual images and visual percepts are alike in many respects.[1] This apparent similarity has led to a great deal of research assessing the parallels between images and percepts by searching for common structural properties of percept-based and image-based representations as well as shared neurological substrates between the visual perception and imagery systems. For example, considerable research has demonstrated that both images and percepts encode spatial dimensions and spatial relations among objects (e.g., Farah, 1985; Finke & Pinker, 1982; Kosslyn, 1973, 1978; Kosslyn, Ball, & Reiser, 1978; Pinker, 1980; Pinker & Kosslyn, 1978; Weber & Harnish, 1974). In addition, decision times for comparison of two imagined objects correspond to those for two perceived objects (e.g., Moyer & Bayer, 1976; Paivio, 1975). Like visual perception, visual images are limited in spatial resolution in that the amount of information contained in an image is constrained by the imagined object's size and distance (e.g., Kosslyn, 1975, 1976, 1978; Weber & Malmstrom, 1979) as well as eccentricity in the imagined visual field (e.g., Finke & Kosslyn, 1980; Finke & Kurtzman, 1981b). Furthermore, just as closer visual inspection of an object from different points of view reveals previously unrecognized properties of the object, there is some evidence that manipulation of mental images can also reveal patterns in the image not apparent when the image was originally formed (e.g., Brandimonte, Hitch, & Bishop, 1992; Pinker & Finke, 1980), although it appears that such insight is more difficult with images than with percepts (e.g., Chambers & Reisberg, 1985).

James R. Brockmole, Ranxiao Frances Wang, and David E. Irwin, Department of Psychology, University of Illinois at Urbana–Champaign.

This research was conducted by James R. Brockmole in partial fulfillment of the requirements for the degree of master of arts at the University of Illinois at Urbana–Champaign. Support for this research was provided by National Science Foundation Grant SBR 96-15988 to David E. Irwin, a University of Illinois Institutional Research Board grant to Ranxiao Frances Wang, and a National Science Foundation graduate research fellowship to James R. Brockmole.

We thank Vincent Di Lollo, Thomas Sanocki, and Zenon Pylyshyn for their helpful comments, as well as Elizabeth Hennesey and Renee Mehl for their help with data collection.

Correspondence concerning this article should be addressed to James R. Brockmole, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, Illinois 61820. E-mail: jbrockmo@s.psych.uiuc.edu

The behavioral findings that suggest that image-based and percept-based representations share a common structure are complemented by research that has shown that visual imagery relies upon some of the same neural machinery as visual perception and generates similar patterns of neural activation.[2] For example, using a radioactive marker to observe cerebral blood flow, Roland and Friberg (1985) determined that imagining an object is associated with activation in the occipital, posterior superior parietal, and posterior inferior temporal lobes, all of which are involved in perceptual processing (see also Goldenberg et al., 1989). In addition, using positron emission tomography, Kosslyn and his colleagues (Kosslyn et al., 1993) showed that the locus of activation in the primary visual cortex (occipital lobe) during image generation depends upon the size of the imagined object. The primary visual cortex is topographically mapped in accordance with input received from the retina—information received from the fovea causes activation posteriorly, whereas information from the periphery causes activation in progressively more anterior areas. Accordingly, imagining small objects (which, when seen, would mostly occupy the center of the field of view) was associated with posterior activation, whereas imagining large objects (which, when seen, would occupy more peripheral regions of view) was associated with more anterior activation of the occipital lobe. Furthermore, studies of abnormal neurological cases have shown that damage to the primary visual cortex, as well as to the higher-level processing dorsal (posterior parietal lobe) and ventral (posterior temporal lobe) visual systems can result in impairments in imagery use (e.g., Farah, Soso, & Dasheiff, 1992).

[1] The nature of visual images has been a debated topic in the literature. A specific question of controversy is whether images are analog or propositional in form. The same debate, however, extends to the nature of representations derived directly from percepts (vision proper). This distinction is not important to the research reported here, because the impetus for our experiments rests upon the similarities between percepts and images. Whether, in the end, percepts and images are analog or propositional in form is irrelevant to our experiments (see Pylyshyn, 1973, 1978, 1981; for reviews and commentary on this debate, see also Finke, 1980, 1985; Kosslyn, 1997).

[2] For complete reviews, see Farah (1984, 1988), Kosslyn (1987), and Tippett (1992).

On the basis of research that has demonstrated both structural similarities in the mental representations formed from percepts and images, as well as research showing imagery and perception share neurological substrates, Kosslyn (e.g., 1980, 1995, 1997; Kosslyn & Koenig, 1992) has suggested that visual imagery and visual perception are supported by the same cognitive architecture (see also Shepard, 1984). Kosslyn argues that within the visual system, there is a visual buffer that maintains currently activated visual representations. Information can enter the buffer from either direct sensory input (perception) or from memory (imagery). Furthermore, once information is loaded into the buffer, it is processed equivalently, regardless of its point of origin. According to Kosslyn, then, an image is a short-term visual representation that represents surface and spatial layout information in an arraylike format in the absence of immediate sensory input and is treated by the visual system as though it arose from direct perception (see also Farah, 1985). Because perception- and imagery-based representations have similar properties and are processed by many of the same neurological structures, Kosslyn argues that it should be possible for interactions to occur between images and percepts. In this article we adopt Kosslyn's definition of visual imagery and investigate whether images and percepts can be combined to form new, composite representations.

Previous research has addressed the issue of image–percept interaction by exploring the effect of previously formed images on current visual perception. For example, in one of the earliest demonstrations of image–percept interaction, Perky (1910) showed that while intentionally imagining an object, subjects were less able to detect a brief visual presentation of that object as compared with cases where imagery was absent (see also Neisser, 1976; Posner, Boies, Eichelman, & Taylor, 1969). More modern replications of the Perky effect have repeatedly shown that visual detection (e.g., Segal, 1971; Segal & Fusella, 1970) and discrimination (e.g., Craver-Lemley & Reeves, 1992; Craver-Lemley, Arterberry, & Reeves, 1997; Reeves, 1981) are impaired by visual imagery (see also Craver-Lemley, Arterberry, & Reeves, 1999). However, imagining a visual scene facilitates object recognition for items that are semantically related to the imagined scene (Peterson & Graham, 1974), and an imagined context also aids in making judgments about visual patterns (Freyd & Finke, 1984).

Whereas previous studies have investigated whether forming an image facilitates or inhibits visual perception, in the present article we attempt to determine whether visual images and visual percepts can be integrated to allow a perceptual judgment involving the combination of the two and to investigate the properties of such an integration process. To do this, we used the temporal integration paradigm originally developed by Eriksen and Collins (1967) to explore the integration of two individual percepts (see also Di Lollo, 1980; Di Lollo & Hogben, 1987; Di Lollo, Hogben, & Dixon, 1994; Dixon & Di Lollo, 1994; Loftus & Irwin, 1998). During the temporal integration procedure, a stimulus is divided into two parts. These parts are presented serially, separated by some time delay. The subject's task requires being able to "see" the whole stimulus, that is, the sum of the two parts. In Eriksen and Collins's original task, two seemingly random dot patterns were individually presented on each trial. However, if superimposed, the patterns created a nonsense, three-letter syllable. The subjects were to identify the syllable, a task that required the combination of information from the two distinct (not temporally or physically overlapping) visual presentations.

Performance in a temporal integration task is generally considered to indicate the degree to which two stimuli are integrated. Studies on temporal integration between two stimuli have shown that performance is very high when the temporal separation between the two stimulus parts is very short (e.g., less than 50 ms). However, increasing the stimulus onset asynchrony (SOA) between the two stimulus parts by increasing either the presentation duration of the first stimulus or the interstimulus interval (ISI) between the offset of the first stimulus and the onset of the second stimulus profoundly hinders performance (Di Lollo, 1980; Loftus & Irwin, 1998). When the SOA is greater than 50 ms, subjects' performance decreases dramatically and reaches the floor at about 100 ms. These results suggest that the combination of two visual stimuli is possible, but only if they fall in a very small temporal window. This is because visible persistence from the first stimulus must be integrated with the presentation of the second stimulus, and visible persistence is short-lived and time-locked to stimulus onset (Coltheart, 1980; Di Lollo, 1980).

To test whether integration can also occur between visual images and visual percepts, we extended the temporal delay between the two stimulus parts long enough so that a visual image of the first stimulus could be formed. If that image can be integrated with the second stimulus, then performance should improve, and systematic variation of the temporal delay should allow us to measure the characteristics of the image formation process and the integration process.

Experiment 1

To test the image–percept integration hypothesis, Experiment 1 used a variant of the temporal integration task devised by Di Lollo (1980) in which subjects made responses based upon the combination of information from two discretely presented stimuli. Subjects were shown an enclosed 4×4 square grid. On any given trial, several individual squares were filled simultaneously with a dot for a brief time. After some delay (during which the grid was blank), several previously empty squares were each briefly filled with a dot. In the course of this process, one square in the grid was never filled. The subject's task was to indicate which position was left empty. To accomplish this goal, subjects had to have knowledge of the positions of all the dots that were presented.

The delay separating the two dot-array presentations was manipulated by varying the ISI between the arrays. ISIs ranged from 0 to 5,000 ms in intervals of 500 ms, although an ISI of 100 ms was also included. ISIs of 0 and 100 ms were used to replicate the general findings associated with perceptual integration and to serve as baseline conditions. Performance should be at ceiling when no delay separates the dot arrays, because the two arrays should be perceived as a single presentation. At the other extreme, performance should fall to floor when 100 ms separates the array presentations. The longer ISIs were used to test the hypothesized *image–percept integration process* that may occur beyond the temporal window for perceptual integration. The image–percept integration hypothesis posits that when subjects are afforded enough time to form an image of the first dot array, performance should improve as ISIs increase above 100 ms, although the level of performance that is ultimately reached may or may not be equivalent to the ceiling level established when the two arrays were integrated perceptually.

Method

Subjects

Eight members of the University of Illinois community, including two of the authors (J.R.B. and R.F.W.), participated after providing informed consent. Except for the authors, all subjects were naive with respect to the experimental hypotheses and were paid $5 for participating.

Stimuli

Stimuli consisted of two unique dot arrays presented within an enclosed 4×4 square grid. The first array contained seven dots and the second array contained eight dots. Together, the arrays filled all but one square in the grid. The grid was composed of interconnected lines, drawn over the background such that the color within the grid spaces and the area surrounding the grid was the same. Subjects viewed the stimuli at a normal viewing distance from a computer screen (approximately 50 cm). The total display subtended 38.1 cm (approximately 37° of visual angle) horizontally and 27.9 cm (29°) vertically. The square grid subtended 15.9 cm (18°). Each square within the grid subtended 3.96 cm (4.5°). Each dot presented in the array subtended 3.3 cm (3.8°). The display background was light gray, the grid lines were light blue, and the dots were black.

Apparatus

The stimuli were presented at a refresh rate of 60 Hz on a Gateway VX900 monitor. A Gateway E-4200 microcomputer controlled stimulus presentation and recorded responses. Subjects made their responses by moving a cursor controlled by a mouse trackball and clicking on the grid location that they thought had been unfilled.

Design and Procedure

Each trial consisted of two sequentially presented dot arrays within an enclosed square grid separated by a variable ISI. On any given trial, one position within the grid was never filled. Under the constraint that no members of the two arrays occupy the same spatial position, the dot patterns were randomly generated. Accordingly, the location of the missing dot was equally likely to occur in all positions within the grid.

Subjects were instructed that they were to identify the position of the missing dot. They were told that in order to do this they would have to remember the positions of the dots in Array 1 across the delay that separated the arrays and combine this information with that contained in Array 2. It was suggested that a good strategy to use would be to imagine the dots still being present after they disappeared.

Prior to the onset of each trial, an empty 4×4 square grid was presented in the center of the display. When ready to begin the trial, subjects pressed the spacebar. Immediately, the first array of seven dots was presented within the grid for 33 ms. The ISI between the offset of the first dot array and the onset of the second array was either 0, 100, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, or 5,000 ms, during which the empty grid remained on the screen. Following the delay, the second array containing eight dots was presented for 33 ms. After both arrays were presented, a cursor appeared on the screen. Using a mouse, subjects moved the cursor to the position in the grid that they thought was left empty during the trial and clicked one of the buttons. Subjects were told to respond as accurately as possible and that they were under no speed stress.

Each ISI (12) occurred equally often. Using a repetition factor of 32, each subject completed a total of 384 trials, divided into blocks of 8 trials. Within each block, the ISI was constant. Subjects were informed of the ISI duration prior to the start of each block. For each subject, the blocks of trials were presented in a different random order. The subjects' responses were recorded.

Prior to beginning the experimental trials, subjects completed 25 practice trials. Practice trials consisted of five trials each, at ISIs of 0, 100, 1,000, 2,000, and 3,000 ms. During practice, subjects were given feedback concerning the accuracy of their responses. However, during the experimental trials no feedback was given.

Results and Discussion

The results are reported in four sections. First, as a preliminary measure of performance, accuracy was assessed as a function of ISI. Second, because accuracy is affected by the amount of information retained from either array, the source of errors was assessed with an emphasis on the strength of the relationship between different types of error and accuracy. Third, the number of dots remembered from each array was calculated to quantify the amount of available information that was integrated. Finally, trends in performance were assessed in order to address the temporal characteristics of the underlying processes.[3] Trends in performance, as measured by accuracy, error, and capacity, are illustrated in Figure 1. For all analyses, an alpha level of less than .05 was adopted as the criterion for statistical reliability.

Accuracy

According to the image–percept integration hypothesis, visual images and visual percepts are integrated in the visual buffer. Thus, as more information from the first array enters the visual buffer and becomes integrated with the second array, accuracy should increase. Accuracy was measured in terms of the percentage of trials on which subjects correctly identified the location of the missing dot. Analysis of accuracy focused on the effect of increasing the ISI between dot array presentations.

A one-way repeated measures analysis of variance (ANOVA) with ISI as the independent factor demonstrated that accuracy did vary as a function of ISI, $F(11, 77) = 22.1$, $MSE = .930$. Single-degree-of-freedom polynomial tests showed a reliable cubic trend between ISI and accuracy, $F(1, 7) = 49.3$, $MSE = .803$, indicating nonmonotonicity in the accuracy data. Visual inspection of mean accuracy as a function of ISI suggested that two distinct monotonic functions represented the data. First, accuracy decreased as the ISI increased from 0 to 100 ms. Second, there was a curvilinear increase in accuracy between ISIs of 100 and 5,000 ms. Each of these trends is discussed in turn.

Accuracy when the ISI was less than or equal to 100 ms (perceptual integration). When the ISI between array presentations was 0 ms, accuracy for identifying the missing dot was very good, averaging 79%. However, when the ISI increased to 100 ms, accuracy dropped to a mean of 21%. This difference in performance was reliable, $t(7) = 12.0$. On the basis of previous perceptual integration research and subjective reports obtained from subjects in this study during debriefing, the two dot arrays seemed

[3] Each measure of performance is not completely orthogonal and some redundancy does exist. However, each measure characterizes performance in a different light, and therefore each contributes to an overall depiction of the phenomenon being studied.

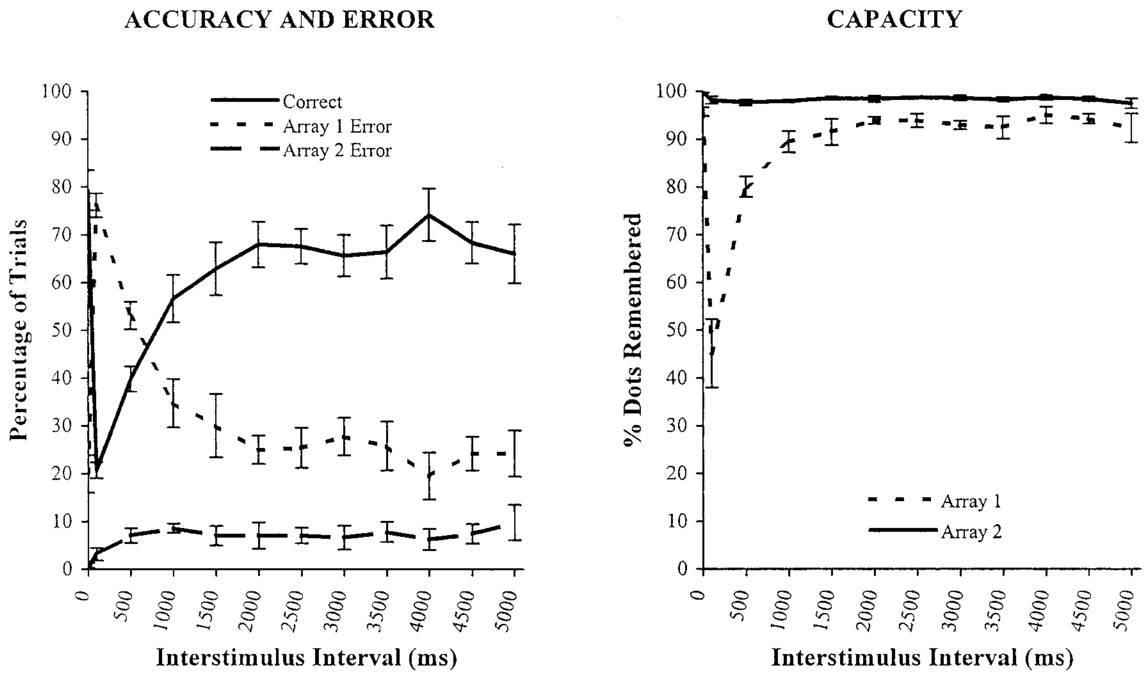

Figure 1. Results of Experiment 1. Left: Mean correct responses, Array 1 errors, and Array 2 errors (with standard error bars) as the percentage of trials on which they occur as a function of interstimulus interval (ISI). Right: Percentage of dots from each array (with standard error bars) that was remembered as a function of ISI.

to be perceptually integrated into a single presentation when no delay separated them, although the arrays were never physically present concurrently. However, when the delay between array presentations was 100 ms, the integration effect was greatly reduced and performance decreased sharply. Consistent with previous research, these data showed that integration between two percepts can occur, but only within a very small temporal window ($<$ 100 ms).

Accuracy when the ISI was greater than or equal to 100 ms. When the ISI was 100 ms or greater, accuracy also varied as a function of ISI, $F(10, 70) = 22.2$, $MSE = .890$. Single-degree-of-freedom polynomial tests demonstrated a reliable positive-slope linear trend, $F(1, 7) = 68.9$, $MSE = 1.67$, and quadratic trend, $F(1, 7) = 75.5$, $MSE = .860$, in the accuracy data. The reliable positive-slope linear trend indicates that accuracy increased as ISI increased. The reliable quadratic trend indicates that the rate of increase in accuracy was not constant across all ISIs. That is, the relationship between accuracy and ISI was curvilinear. Post hoc contrasts showed that accuracy improved as the ISI increased through 1,500 ms but did not differ when the ISI was greater than or equal to 1,500 ms. After 1,500 ms, accuracy asymptoted at approximately 68%. It is important to note that no decrease in accuracy was detected at later ISIs.

Errors

Accuracy is affected by the amount of information retained from both the first and second array. Thus, from the accuracy data alone, the source of improvement is not clear. When an error is made, it could be of two forms. First, the incorrectly selected position of the missing dot could have been previously occupied by a dot from the first array; this kind of error results from a failure to properly retain or use information from the first array. Second, the incorrectly selected position could have been previously occupied by a dot from the second array; this kind of error is associated with a failure to perceive or encode information contained in the second array, possibly due to the interference of the first array. Each type of error was measured in terms of the percentage of trials on which they occurred.

As a percentage of total trials, Array 1 errors occurred, on average, on 32% of the trials and accounted for 81% of all errors. The Pearson correlation between Array 1 errors and correct responses across ISIs was $-.95$, revealing an almost perfect dependence between accuracy and Array 1 error. On the other hand, as a percentage of total trials, Array 2 errors occurred, on average, on 6% of the trials and accounted for 19% of all errors. Array 2 errors were rare ($<$ 1%) when the ISI was 0 ms; when the ISI exceeded 0 ms, Array 2 errors occurred on 7% of trials and did not vary with ISI. The Pearson correlation between Array 2 error and accuracy across ISIs was $-.17$, revealing relative independence between accuracy and Array 2 error (even stronger independence was observed when the ISI exceeded 0 ms, in which case the correlation was $-.09$). Overall, it appears performance depended primarily upon the quality of the representation formed or maintained of the first array.

Capacity

Although accuracy provides an initial test of the image–percept integration hypothesis, it can be misleading as an estimate of how much information was used in the judgment task because it is not necessarily a direct reflection of the amount of information remembered. Accuracy increases as more and more dots are remembered—from either array. However, such increase is not linear. For instance, forgetting only one dot will cause accuracy to drop from 100% to 50%, assuming that subjects make a random choice between the correct cell and the forgotten one. Forgetting one more dot does not do as much harm, though, as performance will drop to 33% (i.e., choosing 1 of 3). Additionally, error rates can indicate the source of improvement in accuracy but cannot quantify changes in the amount of information that is retained from each array. Thus, a more direct assessment of the amount of information used for the judgment (i.e., the number of dots remembered) is useful in assessing properties of image–percept integration, such as the capacity of the representations formed.

With a few simple assumptions, the amount of information that was retained from each array can be calculated. The amount of information retained from the two arrays can be determined by combining both accuracy and error information. The percentage of Array 1 errors to correct responses is equivalent to the number of dots forgotten from the first array. Similarly, the percentage of Array 2 errors to correct responses is equivalent to the number of dots forgotten from the second array (see Appendix A for a detailed discussion of the logic behind these calculations). Therefore, the number of image and percept dots remembered are calculated by simply subtracting the number of image and percept dots forgotten from the number of dots actually presented in the array.[4]

The first array. On average, 6.2 (88%) of the dots presented in the first array were remembered across the ISI. As would be expected on the basis of trends in accuracy and Array 1 error, the capacity of the first array varied as a function of ISI, $F(11, 77) = 45.4$, $MSE = .352$. Not surprisingly, capacity was greatest when the ISI was 0 ms ($M = 6.7$ dots, or 96%), and lowest when the ISI was 100 ms ($M = 3.0$ dots, or 43%). After 100 ms, a curvilinear increase in capacity was observed. Capacity no longer increased once the ISI reached 1,500 ms, after which capacity attained a mean level of 6.6 dots (94%). Planned comparisons showed that this level of capacity did not reliably differ from that when the ISI was 0 ms, $F(1, 7) = 2.1$.

The second array. The amount of information retained from the second array was very high, averaging 7.8 dots (98%). No effect of ISI was observed, $F(11, 77) = 1.3$, which is consistent with the absence of an effect of ISI on Array 2 error when the ISI was greater than 0 ms. In addition, no reliable interpretable trends were observed, and the Pearson correlation between ISI and capacity in the second array ($-.08$) also failed to reach statistical reliability. The near-perfect capacity for the second array and its invariance across ISIs suggests that little forward masking or interference from the first array occurred in this task.

Temporal Characteristics of Performance

A critical analysis of Experiment 1 concerns the temporal characteristics of improvement, which were measured in terms of the amount of time it takes for performance to reach asymptote. The analyses using ANOVA models (reported above) suggested that accuracy, Array 1 errors, and capacity no longer varied after 1,500 ms. However, a disadvantage to using an ANOVA model to estimate the point at which each measure becomes independent of ISI is that such an estimate is restricted to ISI values that were chosen a priori. The true estimate is likely to fall between the selected ISI values.

This limitation was addressed using linear spline regression analyses. Using this procedure, two contiguous line segments were fit to the performance data by varying the slope of the first segment and the joint between the lines. Because the various ANOVA analyses showed that performance (whether it be measured by accuracy, Array 1 error, or capacity) ultimately reached an asymptote, the second line had a fixed slope of 0 (for a detailed description of this approach, see Appendix B). Because capacity is the best measure of the underlying process by which performance improves, further analysis on the temporal characteristics of performance used capacity as the dependent measure. More specifically, because performance depended primarily upon the amount of information derived from the first array, and because the amount of information retained from the second array remained stable across all ISIs, the capacity of the first array was analyzed.[5]

The best fitting lines to the capacity data showed an increase in capacity for the first array through 1,300 ms, after which capacity reached asymptote. Thus, this analysis suggests that a slow process, such as forming an image of the first array, is responsible for the increase in performance, and it requires approximately 1,300 ms to complete. This estimate is consistent with other two-dimensional image generation studies, which have generally estimated that images require between 1,000 and 2,000 ms to generate (e.g., Kosslyn, Cave, Provost, & von Gierke, 1988; Kosslyn, Reiser, Farah, & Fliegel, 1985; Weber & Harnish, 1974).

There is, however, an alternative possibility. The improvement in performance as ISI increases may potentially reflect the dissipation of backward masking, rather than the development of a visual image. Studies of perceptual integration have shown that backward masking occurs between two successively presented visual arrays, whereby the second array limits the amount of information that can be extracted from the leading array (e.g., Di Lollo & Hogben, 1987; Groner, Groner, Bischof, & Di Lollo, 1990). Thus, more information about the first array might be remembered with longer ISIs because there is less masking between the two presentations when the temporal separation increases. The rather long time course of improvement (1,300 ms) indicates that backward masking alone cannot explain the improvement in performance in the task, however. The effects of backward masking should have dissipated after 300 ms (Averbach & Coriell, 1961; Breitmeyer, 1980, 1984; Breitmeyer & Ganz, 1976; Kahneman, 1968; Lefton, 1973; Matin, 1975; Turvey, 1973). However, capacity continued to improve for at least 1,000 ms beyond this time. These results contradict a pure backward masking interpretation (although masking might have hindered performance in the 100-ms ISI condition) and provide further support for the hypothesis that, when given sufficient time, an

[4] By number of dots remembered, we do not mean that each dot was remembered as a separate entity; rather, some chunking may have taken place. Therefore, this analysis may have inflated the underlying capacity for individual entities. It nonetheless provides a measure of the amount of information from the array that was remembered.

[5] The same analysis was also performed on trends in accuracy and error and produced results consistent with those using capacity as the dependent measure.

image of the first array was generated and combined with ongoing perception.

Although conventional mechanisms of backward masking cannot account for the present data, it is possible that presentation of Array 2 interfered with the continuing development of Array 1 when the ISI was less than 1,500 ms. That is, once the second array was presented, it appears that subjects were unable to maintain and develop the image of Array 1 further so that it could be integrated with the slowly developing image of Array 2. There may be several reasons for the cessation of image generation once the second array was presented. One possibility is that an attentional switch away from image generation of Array 1 may have been triggered by the onset of the second array.[6] This is possible within Kosslyn's (1980, 1995, 1997) conception of the visual buffer, because it contains an *attention window* that selects input on the buffer for further processing. If the second array draws this window away from the forming image, further generation may not be possible. Another possibility is that once the second array is presented and integrated with the information from the first array that was already loaded into the visual buffer, continued generation of the first array would require generating and maintaining a representation of the positions of all 15 presented dots—a process that might conceivably exceed limitations on short-term memory or image capacity. Issues of integration speed and capacity limitations are considered in Experiments 2–5.

Summary

The results of Experiment 1 are consistent with the image–percept integration hypothesis. When no temporal delay separated the presentation of the two dot arrays, accuracy was very high (79%). However, when 100 ms separated the array presentations, accuracy declined to 21%. When the ISI exceeded 100 ms, accuracy improved to approximately 68%. Error analysis showed that this increase in accuracy was almost exclusively associated with a decrease in Array 1 errors, and capacity analyses showed that the amount of information retained from the first array rose from 43% (i.e., about 3.0 dots) when the ISI was 100 ms to a ceiling level of approximately 94% (i.e., about 6.6 dots). These results demonstrate that as the ISI increased beyond 100 ms, subjects were better able to represent information from the leading display.

Capacity for the first array increased at a relatively slow and constant rate and did not reach asymptote until 1,300 ms. This temporal characteristic of capacity is inconsistent with the hypothesis that accuracy improves simply because backward masking is removed at longer ISIs. Under such a hypothesis, capacity should have recovered more quickly (certainly by 500 ms), according to previous research on masking. Rather, a slow and effortful process that enabled subjects to represent more and more information from the first array long after its disappearance seems to underlie performance. Thus, it appears that subjects may generate an image of the first array during the ISI that can subsequently be integrated with incoming visual information from the environment. Finally, capacity for the first array for ISIs greater than or equal to 1,500 ms did not differ from capacity when the ISI was 0 ms (i.e., during perceptual integration). Therefore, the process guiding integration at long ISIs, in this experiment, produces a representation almost as rich in information as that following perceptual integration.

Experiment 2

Experiment 1 demonstrated that performance in a temporal integration task recovered slowly after the dissipation of perceptual integration and reached asymptote after 1,300 ms. These results support the hypothesis that a visual image of the first array was generated during the delay, which was subsequently integrated with a percept. Experiment 2 further tested whether the integration process was between an image and a percept or between two images by measuring how much time such an integration process takes in comparison with the perceptual integration process.

It is important to note that this question is very different from questions concerning the temporal characteristics of performance explored in Experiment 1. Experiment 1 addressed how much time is needed to create an adequate representation of the first array in the visual buffer; Experiment 2 addresses how much time is needed to process the second array and subsequently integrate the arrays to arrive at a judgment (i.e., after the first array is already in the visual buffer). According to the image–percept integration hypothesis, the time it takes to process the second array and integrate the two representations should be similar during image–percept integration and perceptual integration, although the decision time might be slightly different, depending on the number of alternatives from which the subject must choose. If, contrary to our hypothesis, subjects formed separate images of both arrays, which were then combined into a unified representation, then the process that governs performance at late ISIs is not the direct integration of images and percepts, but rather the interaction of two memory traces. Such a process should require much more time to complete than a process that directly integrates two percepts or a percept with the image in the visual buffer because time must be spent generating an additional image (somewhere on the order of 1,300 ms).

To address the time required to complete the integration and judgment process, we carried out a variant of Experiment 1 where subjects' reaction time (RT) was recorded. Aside from recording RTs, two important changes were introduced. First, only two ISIs were used. Specifically, the ISI between arrays was either 0 ms or 2,000 ms. When the ISI is 0 ms, the two dot arrays are perceptually integrated. When the ISI is 2,000 ms, perceptual integration is no longer possible and the image–percept integration process would govern performance. Note that 2,000 ms exceeds the 1,300 ms estimate in Experiment 1 for how much time is required for performance to reach ceiling once perceptual integration is no longer possible. Second, the time between the offset of the second array and the onset of a cursor, which subjects moved in order to register their responses, was systematically varied. Because subjects could not overtly respond until the cursor appeared, this manipulation had the effect of delaying subjects' ability to register their judgment after the second array was presented for some time. Thus, we refer to the time between the offset of the second array and the onset of the cursor as the *response delay*. The response delay manipulation is critical in estimating the time required to integrate the two arrays and form a judgment in the two types of integration processes.

[6] We thank Vincent Di Lollo for suggesting this possibility; see Di Lollo and Moscovitch (1983).

At short response delays, the integration and judgment processes do not have time to be completed prior to the presentation of the cursor (when RT starts to be timed). Therefore, RT will reflect array integration time, judgment time, and motor response time. However, as the response delay increases, a greater portion of the integration–judgment process is completed prior to the presentation of the cursor and would therefore not be included in the RT measure. This results in a decrease in RT as response delay increases. At some point, the cursor delay will exceed the time required to integrate the arrays and form a judgment. When this point is reached, RT will include only the time to complete the motor response, and the RT curve will asymptote at some floor level. The point at which this occurs provides a measure of how much time is required to integrate the two arrays and complete the cognitive process of selecting the location of the missing dot.

Method

Subjects

Eight members of the University of Illinois community participated after providing informed consent. None of the subjects in Experiment 2 participated in Experiment 1. All subjects were naive with respect to the experimental hypotheses and were paid $10 for participating.

Stimuli

The same stimuli were used as in Experiment 1.

Apparatus

The same apparatus was used as in Experiment 1.

Design and Procedure

As in Experiment 1, each trial consisted of two sequentially presented dot arrays within an enclosed square grid separated by some delay (ISI). On any given trial, one position within the grid was never filled; the subject's task was to identify the position of the missing dot. Prior to the onset of each trial, an empty 4×4 square grid was presented in the center of the display. When they were ready to begin the trial, subjects pressed the spacebar. Immediately, the first array of seven dots was presented within the grid for 33 ms. In separate blocks, the ISI between the offset of the first dot array and the onset of the second array was either 0 ms or 2,000 ms. Following the delay, the second array containing eight dots was presented for 33 ms. After both arrays were presented, a cursor appeared on the screen in a random location. However, the elapsed time between the offset of the second array and the appearance of the cursor varied. The cursor appeared either 0, 67, 167, 333, 467, 600, 767, 900, 1,067, 1,200, 1,367, or 2,000 ms after the offset of the second array. Subjects could not register a response until the cursor appeared. Thus, this manipulation is referred to as the response delay. Using a mouse, subjects moved the cursor to the position in the grid that they thought was left empty during the trial and clicked one of the buttons. Subjects were told that their reaction time was being recorded. However, to minimize a speed–accuracy tradeoff, subjects were not told to respond as quickly as possible but rather to respond as soon as they knew the answer.

Response delay and ISI were crossed. Each response delay (12) occurred equally often. Using a repetition factor of 32, each subject completed 384 trials at each ISI, or 768 total trials. Trials were divided into four sessions, each consisting of 192 trials. Each session was further divided into 24 blocks of 8 trials. Within each session the ISI was constant; within each block the response delay was constant. Subjects were informed of the duration of both the ISI and response delay prior to the start of each block. For each subject, the blocks of trials were presented in a different random order. An ABBA counterbalance was used to determine the order of sessions. Sessions were completed on 2 separate days. On each day subjects completed two sessions at different ISIs. The subjects' responses and reaction times were recorded.

At the beginning of each day, before the experimental trials, subjects completed 100 practice trials. Practice trials consisted of 10 trials, each using response delays of 0, 500, 1,000, 1,500, and 2,000 ms at each ISI. During practice, subjects were given feedback concerning the accuracy of their responses. However, during the experimental trials no feedback was given.

Results and Discussion

The purpose of Experiment 2 was to obtain an estimate of the time required to integrate the two arrays and arrive at a judgment when either 0 ms (perceptual integration) or 2,000 ms (hypothetical image–percept integration) intervened between array presentations. Analyses therefore focused on RT as a function of response delay. However, prior to such an analysis it was necessary to ensure that other dimensions of performance aside from RT did not vary as a function of response delay.

Accuracy, Error, and Capacity

Separate, one-way, repeated measures ANOVAs were conducted on accuracy, Array 1 error, Array 2 error, and capacity for each array to explore potential differences across response delay within each ISI. In all analyses, none of these measures of performance varied as a function of response delay. On average, when the ISI was 0 ms, accuracy was 86% ($SEM = 0.5$); Array 1 error occurred on 12% of trials ($SEM = 0.4$); Array 2 errors occurred on 2% of trials ($SEM = 0.3$); 6.8 dots ($SEM = 0.01$), or 97%, were remembered from the first array; 7.9 dots ($SEM = 0.01$), or 99%, were remembered from the second array. When the ISI was 2,000 ms, accuracy was 53% ($SEM = 1.0$); Array 1 error occurred on 40% of trials ($SEM = 1.2$); Array 2 error occurred on 7% of trials ($SEM = 0.4$); 6.1 dots ($SEM = .04$), or 87%, were remembered from the first array; 7.8 dots ($SEM = 0.01$), or 98%, were remembered from the second array. Comparisons across ISI indicated reliable differences in accuracy, Array 1 error, Array 2 error, and capacity that mirrored similar comparisons in Experiment 1.

Temporal Characteristics

Having established that accuracy, error, and capacity did not vary by response delay, the effect of response delay on RT was analyzed. The analysis of RT focused on trials on which a correct response was given. On incorrect trials, there could have been a failure to integrate the arrays completely, and therefore reaction times on these trials may not accurately reflect the time required to integrate the arrays. On the basis of the analysis of accuracy and error as a function of response delay, this trim did not disproportionately exclude trials for either error type from each response delay. Mean RT by ISI and response delay are illustrated in Figure 2. Separate, one-way ANOVAs explored the effect of response delay on RT within each ISI. Each of these analyses is discussed in turn.

RT when the ISI was equal to 0 ms. When the ISI was 0 ms, RT did vary by response delay, $F(11, 77) = 37.5$, $MSE = 4{,}014$,

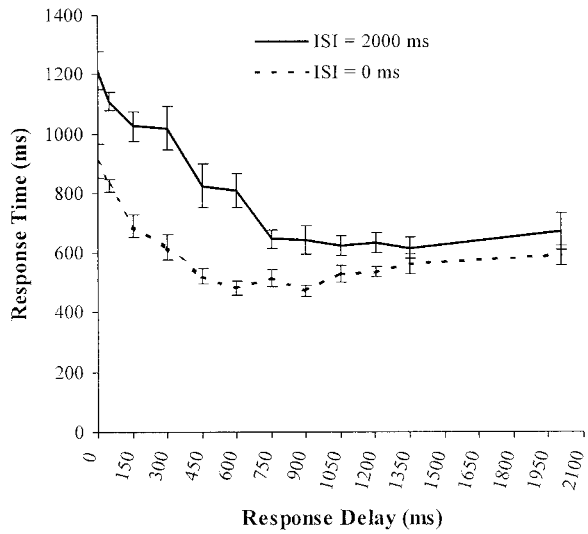

Figure 2. Results of Experiment 2. Mean response times and standard errors by interstimulus interval (ISI) and response delay.

and single-degree-of-freedom polynomial tests showed a reliable negative-slope linear trend in the RT data, $F(1, 7) = 55.0$, $MSE = 14{,}255$. A reliable quadratic trend, $F(1, 7) = 173.4$, $MSE = 4{,}709$, indicated that the decrease in RT as response delay increased did not remain constant across all response delays; the relationship between RT and response delay was curvilinear.

As in Experiment 1, linear spline regression was used to analyze the temporal properties of this relationship. In particular, the point at which RT reached floor was of interest. Pairwise comparisons showed that when response delay was less than or equal to 600 ms, RT reliably decreased; when response delay was between 600 and 1,067 ms, RT reliably increased; when the ISI was at least 1,067 ms, no reliable differences among RTs were observed. Therefore, the second segment of the overall piecewise function in the spline regression analysis was not defined to be a horizontal line. Rather, the slope of both the first and second line segments were free parameters (see Appendix B for further details of this modification). Results of this procedure showed that RT decreased at a rate of 86 ms for every 100-ms increase in response delay through 500 ms, after which RT increased at a rate of 7 ms for every 100-ms increase in response delay. The overall fit of this model was very good, as indicated by a multiple *R* of .82. Simple linear regression of RT on response delay for response delays greater than 500 ms failed to demonstrate a reliable nonzero slope line that best fit the RT data. This suggests RT reached floor at approximately 500 ms.

RT when the ISI was equal to 2,000 ms. When the ISI was 2,000 ms, RT also varied by response delay, $F(11, 77) = 27.5$, $MSE = 13{,}977$, and single-degree-of-freedom polynomial tests showed a reliable negative-slope linear trend in the RT data, $F(1, 7) = 344.4$, $MSE = 10{,}390$. A reliable quadratic trend, $F(1, 7) = 41.2$, $MSE = 10{,}951$, indicated that the decrease in RT as response delay increased did not remain constant across all response delays; the relationship between RT and response delay was curvilinear.

Planned comparisons indicated that RT no longer differed by response delay when response delay was greater than or equal to 767 ms. To be commensurate with the linear spline regression approach used to analyze trials on which the ISI was 0 ms, the slope of both the first and second line segments were free parameters (the second segment was not defined as a horizontal line). This procedure showed that RT decreased at a rate of 70 ms for every 100-ms increase in response delay through 800 ms, after which RT increased at a rate of 1 ms for every 100-ms increase in response delay. The overall fit of this model was also very good, as indicated by a multiple *R* of .83. Again, simple linear regression of RT on response delay for response delays greater than 800 ms failed to demonstrate a reliable nonzero slope line that best fit the RT data. This suggests RT reached floor at approximately 800 ms.

Effect of ISI on integration and processing time. When the ISI was 0 ms, integration and response selection required approximately 500 ms. When the ISI was 2,000 ms, integration and response selection required approximately 800 ms. These estimates were contrasted by computing the time required for RT to reach floor for each subject. The *t* tests showed a reliable difference between the estimates attained from each ISI, $t(7) = 2.3$, indicating that integration and response selection were approximately 300 ms faster when the delay between the arrays was 0 ms than when the delay was 2,000 ms.

Because integration time and response selection time cannot be decoupled, at least in the present paradigm, it is difficult to obtain a precise estimate of how long perceptual integration takes as opposed to response selection. However, it can be concluded that the integration component requires less than 500 ms (the integration plus selection estimate). Likewise, although a specific estimate for how much time image–percept integration takes is equally unattainable, it is no more than 300 ms longer than the perceptual integration process. In other words, even if this slowdown was entirely due to differences in the speed of integration processes, the difference was far less than the time that would be required if an image of the second array were generated prior to selecting a response (1,300 ms, on the basis of the results of Experiment 1). Thus, it seems that integration following a long delay is indeed a product of a relatively direct combination of perceptual information with a previously constructed image maintained in the visual buffer.

Before leaving this discussion, some speculation can be made as to what may be causing the slowdown in the integration–response selection process. If the slowdown is due to differences in the speed of integration, two possible explanations exist. First, perceptual integration may take place earlier in the visual system (e.g., in the retina) than image–percept integration (e.g., in the visual cortex). Second, integration may be slower when memory is required to maintain an image of the first array in the visual buffer. Alternatively, the slowdown may be due to a slower response selection process. In this case, response selection may be slower at the longer ISI because memory is less than perfect, and more time is spent analyzing the integrated representation of the arrays to choose a response from the increased number of alternatives. Additionally, response selection may be a stage of processing that requires access to a limited-capacity bottleneck. A slowdown in response could then arise because they draw upon similar resources as image generation and integration.

Experiment 3

Experiments 1 and 2 have provided good support for the hypothesis that images and percepts can be integrated to form a new, composite representation. Experiments 3, 4, and 5 investigated some of the fundamental properties of this integration process.

The purpose of Experiment 3 was to assess the impact of increasing the duration of the first array on image–percept integration. In Experiments 1 and 2 this duration was fixed at 33 ms, which might have limited the amount of information that subjects could extract from the stimulus in order to form an image of it. Intuitively, it would seem that increasing exposure to the first array would facilitate image generation, because subjects are given more time to transform that array into an image. On the other hand, previous research has shown that increasing the duration of the leading array in a temporal integration task disrupts perceptual integration (e.g., Di Lollo, 1980), so perhaps the same would be true for the integration of images and percepts. For example, longer exposure (i.e., leaving the perceptual stimulus on the screen) could potentially interfere with image generation and maintenance, thereby disrupting the image–percept integration process.

To examine these possibilities, in Experiment 3 the duration of the first array, as well as the delay between array presentations, was varied. That is, the temporal separation of the two arrays was defined by the SOA—the elapsed time between the onset of each array. The SOA was determined either by the delay between the arrays (ISI condition), as in Experiments 1 and 2, or by the duration of the first array (duration condition). The main question addressed in Experiment 3 is whether increasing the duration of the first array will cause more or less improvement in performance as compared with increasing ISI. Such a result would occur if longer experience with the first array enabled subjects to form a better or worse memory trace (i.e., image) of that array.

Method

Subjects

Eight members of the University of Illinois community participated after providing informed consent. None of the subjects in Experiment 3 participated in Experiments 1 or 2. All subjects were naive with respect to the experimental hypotheses and were paid $5 for participating.

Stimuli

The same stimuli were used as in Experiment 1.

Apparatus

The same apparatus was used as in Experiments 1 and 2.

Design and Procedure

The general procedure of Experiment 3 was the same as in Experiments 1 and 2. Each trial consisted of two sequentially presented dot arrays within an enclosed 4×4 square grid where one position within the grid was never filled. The subject's task was to identify the position of the missing dot. As in the previous experiments, after both arrays were presented, a cursor appeared on the screen and, using a mouse, subjects moved the cursor to the position in the grid that they thought was left empty during the trial and clicked one of the buttons. Subjects were told to respond as accurately as possible and that they were under no speed stress.

In separate conditions, the SOA was manipulated by either varying the time that elapsed between the arrays (as in Experiments 1 and 2) or the duration of the first array. The SOA was either 33, 133, 300, 533, 1,033, or 2,033 ms. The SOA depended upon one of two factors: the delay between arrays (ISI condition) or the duration of the first array (duration condition). For the ISI condition, the first array was always presented for 33 ms, and the delay between the offset of the first array and the onset of the second array was 0, 100, 266, 500, 1,000, or 2,000 ms. For the duration condition, the array was presented for 33, 133, 300, 533, 1,033, or 2,033 ms, and 0 ms separated the offset of the first array and the onset of the second array. Note that the SOAs in the two conditions were matched. Also notice that trials on which the SOA was 33 ms for both conditions were physically identical because they used the minimum exposure duration of the first array and the minimum delay between the arrays. Because of this redundancy, crossing SOA and condition yielded 11 different trial types.

Each trial type occurred equally often. Using a repetition factor of 32, each subject completed 352 trials, divided into blocks of 8 trials. Within each block, the trial type was constant. Subjects were informed of the duration of the first array and the duration of the delay between arrays prior to the start of each block. For each subject, the blocks of trials were presented in a different random order. Accuracy for identifying the position of the missing dot within a trial was recorded.

Prior to beginning the experimental trials, subjects completed 70 practice trials. Practice trials consisted of 10 trials each at an SOA of 33 ms and at SOAs of 133, 1,033, and 2,033 ms in each condition. During practice, subjects were given feedback concerning the accuracy of their responses. However, during the experimental trials no feedback was given.

Results and Discussion

The primary interest of Experiment 3 is the comparison of performance in the ISI condition versus the duration condition, with matched SOAs. Trials on which the SOA was 33 ms were considered to be members of both conditions, as it used the minimum values for both the duration of the first array (33 ms) and the time between arrays (0 ms). Thus, analyses comparing conditions focused on performance when the SOA exceeded 33 ms. In three sections, differences in accuracy, error, and capacity across these conditions are considered.[7] Trends in performance, as measured by accuracy, error, and capacity, are illustrated in Figure 3.

Accuracy

Analysis of accuracy focused on the effect of increasing the SOA between dot array presentations. As Figure 3 shows, accuracy varied as a function of SOA in the same way as it did in Experiment 1, and this was true regardless of whether SOA was defined by increases in array duration or increases in delay. In both cases, accuracy decreased as SOA increased from 33 to 133 ms, but then increased as SOA increased from 133 to 2,033 ms. Thus, increasing the duration of the first array affected the overall pattern of performance in this task in much the same way as increasing the ISI.

---

[7] Temporal characteristics were not analyzed. Statistically, capacity shared a direct and linear relationship with SOA. Therefore, there was no statistically identifiable point at which capacity no longer increased. The absence of the quadratic trend observed in Experiments 1 and 2 is probably due to the exclusion of SOAs longer than 2,033 ms in Experiment 3. Thus, there was not enough power to statistically observe the curvilinear data pattern. However, it is clear from examining the data visually that such a trend was present.

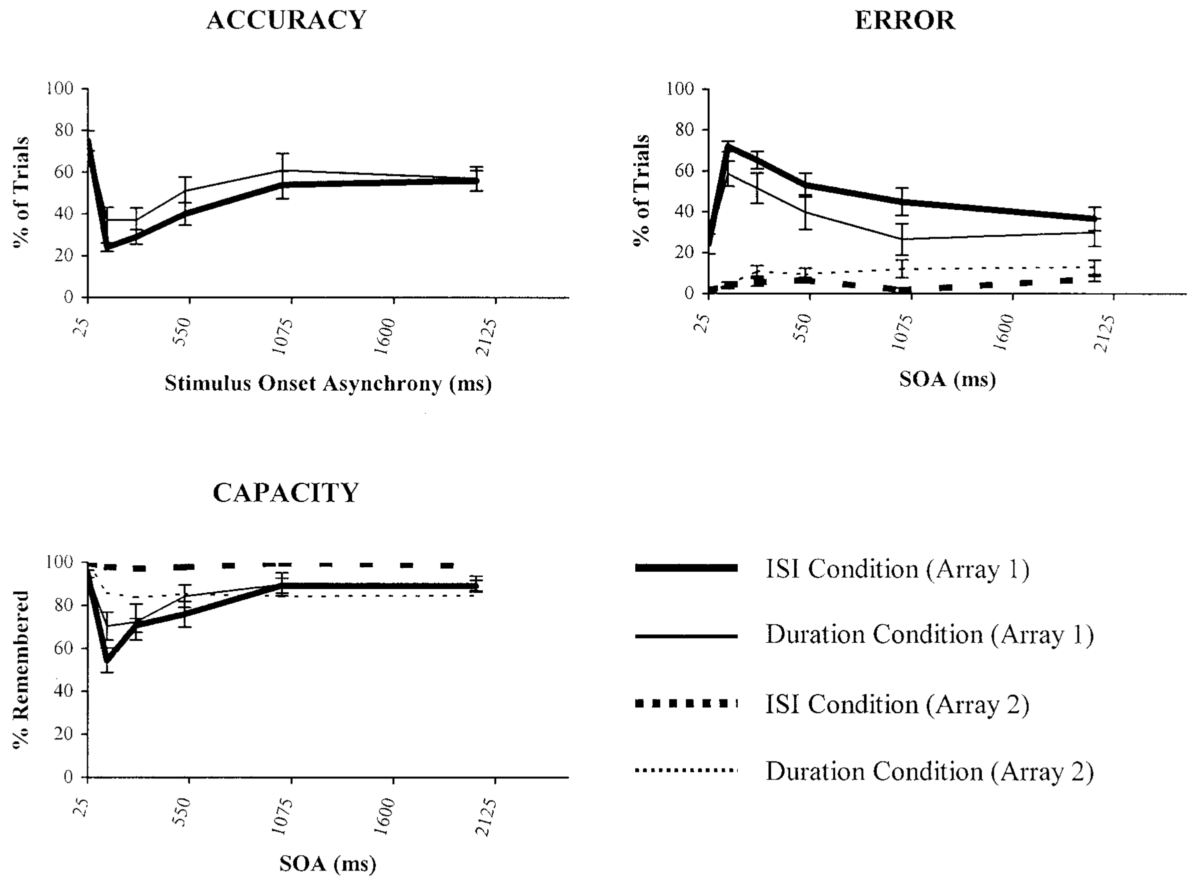

Figure 3. Results of Experiment 3. Mean accuracy (top left), error (top right), and capacity (bottom left), with standard error bars, as a function of stimulus onset asynchrony (SOA) and compared between the interstimulus interval (ISI) and duration conditions. Accuracy is illustrated as a percentage of trials on which a correct response was given. Array 1 error and Array 2 error are also illustrated as the percentage of trials on which they occur. Capacity for each array is illustrated as the percentage of dots remembered. (Error bars are too small to be visible for Array 2 capacity data.)

To examine this more closely, the effect of condition on accuracy was assessed for SOAs greater than or equal to 133 ms using a 2 (condition) × 5 (SOA) repeated measures ANOVA. A main effect of condition was observed, $F(1, 7) = 5.7$, $MSE = .023$. Mean accuracy was higher in the duration condition (49%) than in the ISI condition (41%). The interaction between condition and SOA was not significant, $F(4, 28) = 1.0$.

Error

Array 1 and Array 2 error were measured as the percentage of trials on which each occurred. As in the accuracy data, the overall error patterns are very similar to those of Experiment 1, regardless of whether SOA was defined by increases in stimulus duration or by increases in delay. In both cases, the vast majority of errors were Array 1 errors, and there was a very high negative correlation between Array 1 errors and correct responses but a near-zero correlation between Array 2 errors and correct responses. Thus, as in Experiment 1, integration performance was dependent primarily on the representation of the first array.

To examine this more closely, the effect of condition on error was assessed for SOAs greater than or equal to 133 ms using separate 2 (condition) × 5 (SOA) repeated measures ANOVAs on Array 1 and Array 2 errors. With respect to Array 1 error, a main effect of condition was observed, $F(1, 7) = 9.6$, $MSE = .035$. The occurrence of Array 1 error in the duration condition (41%) was lower than in the ISI condition (54%). Post hoc comparisons demonstrated that differences in Array 1 error across the two delay conditions continued to differ across all SOAs. In addition, condition and SOA did not interact, $F(4, 28) < 1$.

With respect to Array 2 error, a main effect of condition was observed, $F(1, 7) = 6.8$, $MSE = .007$. Mean percentage of Array 2 error in the duration condition (10%) was higher than in the ISI condition (5%). Post hoc comparisons demonstrated that at least marginal differences in Array 2 error between conditions were observed across all SOAs. However, no effect of SOA was observed, $F(4, 28) = 2.0$, nor did SOA interact with condition, $F(4, 28) = 1.5$. This indicates that Array 2 errors, in general, were unaffected by SOA per se; rather, differences in the percentage of Array 2 errors were produced by the condition.

Although accuracy tended to not differ between the two conditions beyond the earliest SOAs, differences in Array 1 and Array 2 errors were observed throughout all SOAs. Array 1 errors were more common in the ISI condition, but Array 2 errors were less common. Thus, longer exposure duration to the first array did seem to decrease the likelihood of error associated with that array, but it caused error associated with the second array to be more common. This trade-off is likely to be the reason for similar levels of accuracy in the two delay conditions.

Capacity

Capacity for the first and second array was calculated in the same manner as in Experiments 1 and 2. As in the analyses of accuracy and errors, capacity varied across SOA in much the same way regardless of whether SOA was defined by stimulus duration or by delay. Capacity was highest when the SOA was 33 ms, was the lowest in both conditions when the SOA was 133 ms, then increased linearly as the SOA increased above 133 ms. The amount of information retained from the second array was very high at all SOAs.

To examine this more closely, the effect of condition on capacity was assessed for SOAs greater than or equal to 133 ms using separate 2 (condition) $\times$ 5 (SOA) repeated measures ANOVAs for the capacity for each array. With respect to the first array, a marginal main effect of condition was observed, $F(1, 7) = 4.0$, $MSE = .020$, $p = .09$, and condition marginally interacted with SOA, $F(4, 28) = 2.1$, $MSE = .003$, $p = .10$. Thus, mean capacity in the duration condition (5.8 dots, 83% of the array) may have been slightly higher than that in the ISI condition (5.5 dots, 79%). However, post hoc comparisons demonstrated that differences in capacity between conditions did not differ when the SOA was greater than 133 ms. This indicates that any advantage provided by the longer duration of the first array was confined to the shortest SOAs, beyond which the representation appeared to have much the same quality regardless of stimulus duration.

With respect to capacity for the second array, a main effect of condition was observed, $F(1, 7) = 539.2$, $MSE = .001$. Mean capacity in the duration condition (7.0 dots, 88%) was lower than that in the ISI condition (7.9 dots, 99%). Post hoc comparisons demonstrated that differences in capacity between the two conditions were observed across all SOAs. However, no effect of SOA was observed, $F(4, 28) < 1$, nor did SOA interact with condition, $F(4, 28) = 1.2$. This indicates that capacity for the second array was unaffected by SOA per se; rather, differences in capacity were produced by the condition.

Summary

At early SOAs, accuracy was better in the duration condition than in the ISI condition, but this advantage seemed to be lost for longer SOAs. Error analyses revealed that increasing the duration of the first array is associated with fewer Array 1 errors, but more Array 2 errors. The most revealing analysis, however, concerned the capacity of information derived from each array. For the capacity of the first array, only a moderate advantage was observed in the duration condition relative to the ISI condition, and this benefit was clearly confined to SOAs of less than 300 ms.[8] The capacity of the second array was always higher in the ISI condition than in the duration condition, possibly because of a forward masking effect.

In short, the results of Experiment 3 indicated that, when given enough time, subjects are able to effectively construct an image of the first array either with or without the presence of the array, and at longer SOAs, there was little difference observed between the two conditions. Thus, in terms of integration, an image of the first array maintained over the blank interval in the absence of perceivable information was as effective as a prolonged presentation of the first array. When SOA was short (< 300 ms), however, a longer exposure to the leading array enabled somewhat better performance. Whether this advantage is directly related to image generation is unclear, however. For instance, the longer duration of the first array may have rendered it more resistant to backward masking.

Experiment 4

The purpose of Experiment 4 was to investigate the characteristics of image–percept integration when the amount of to-be-integrated information is increased. This manipulation allowed two important issues to be addressed. First, in Experiments 1–3, subjects were required to hold seven dots in memory over the duration of the delay, an amount that lies within the limits of a short-term memory store for nonspatially formatted (e.g., verbal) information (e.g., Miller, 1956). As such, it is possible that information from the arrays may not have been represented in an image, or an arraylike format on the visual buffer. Finding that image–percept integration can occur even when the amount of to-be-remembered information exceeds the bounds of stores such as verbal working memory would cast doubt on alternative explanations of our results that are based on nonspatial memory. If imagery underlies performance, the various measures of accuracy, error, capacity, and time course should show the same trends as in the previous experiments.

Previous studies, however, have suggested that images are limited in spatial extent and capacity (Finke & Kosslyn, 1980; Finke & Kurtzman, 1981a; Kosslyn, 1975, 1976, 1978; Weber & Malmstrom, 1979). Thus, increasing the amount of to-be-remembered information could cause an associated decrement in performance (e.g., less remembered information, slower time course, etc). Therefore, a second issue that can be addressed by increasing array size is one of capacity limitations in the image–percept integration process. Such an evaluation was not possible in Experiments 1–3 because, given sufficient time, subjects successfully remembered almost all of the first array, forgetting approximately 10% of the available information, or the equivalent of only 0.75 dots (to equate across experiments, when arrays were separated by 2,000 ms, the percentage of remembered dots was 91% in Experiment 1, 87% in Experiment 2, and 90% in Experiment 3 for both conditions). Indeed, in Experiment 1, capacity for the first array after

[8] Although the capacity of the first array differed between the two conditions only for the shortest SOAs, Array 1 errors were more common in the ISI condition for all SOAs. Because the capacity of the first array directly measures the amount of information from the first array, whereas the Array 1 error measure is mainly, but not strictly, determined by the quality of the image, capacity measures provide the best indicator of the underlying causes of performance.

1,500 ms did not reliably differ from that when the ISI was 0 ms. Thus, to assess both nonspatial memory explanations and capacity issues, the effect of increasing the amount of to-be-remembered information on accuracy, error patterns, and the time course of performance was examined.

Method

Subjects

Eight members of the University of Illinois community, including two of the authors (J.R.B. and R.F.W.), participated after providing informed consent. All subjects participated in Experiment 1 prior to Experiment 4. On average, 111 days elapsed between Experiments 1 and 2. Although subjects were told the purpose of Experiment 1 and the possible outcomes in a debriefing session following that experiment, they were not informed of the results of Experiment 1 until the completion of Experiment 4.

Stimuli

Stimuli consisted of two unique dot arrays presented within an enclosed 5 × 5 square grid. Each array contained 12 dots. Only 1 dot occupied any position within the grid. Subjects viewed the stimuli at a normal viewing distance from a computer screen (approximately 50 cm). The total display subtended 38.1 cm (approximately 37° of visual angle) horizontally and 27.9 cm (29°) vertically. The square grid was the same size as that in Experiment 1, subtending 15.9 cm (18°). Each square within the grid subtended 3.18 cm (3.6°). Each dot presented in the array subtended 2.84 cm (3.2°). The display background was light gray, the grid lines were light blue, and the dots were black.

Apparatus

The apparatus was the same as in Experiment 1.

Design and Procedure

Except for the change in the construction and dimensions of the grid and dot arrays, the design and procedure were the same as in Experiment 1.

Results and Discussion

Of interest in Experiment 4 was the effect of increasing to-be-integrated information on performance. Trends in performance, measured by accuracy, error, and capacity, are illustrated in Figure 4 as a function of ISI. Results of Experiment 4 are considered on their own and in relation to those obtained in Experiment 1 (4 × 4 grid).

Accuracy

The overall pattern of performance was similar to that in Experiments 1–3: Accuracy was high at the shortest delay, dropped precipitously at the 100-ms delay, then gradually increased to an asymptotic level at an ISI of approximately 1,000 ms. Between the ISIs of 0 and 100 ms, accuracy fell from 70% to 11%. It is important to note that accuracy continued to vary when the ISI exceeded 100 ms, $F(10, 70) = 6.1$, MSE = .840. A reliable positive-sloping linear trend $F(1, 7) = 6.9$, $MSE = 1.99$, and quadratic trend, $F(1, 7) = 137.4$, $MSE = .225$, indicated that accuracy and ISI shared a direct but curvilinear relationship. Specifically, accuracy improved until the ISI was greater than or equal to 1,000 ms, after which accuracy reached asymptote at approximately 35%.

The increase in array size had only minor effects on perceptual integration. Although accuracy was reliably lower than in Experiment 1 at 0 ms, $t(7) = 2.4$, as well as at 100 ms, $t(7) = 4.8$, the

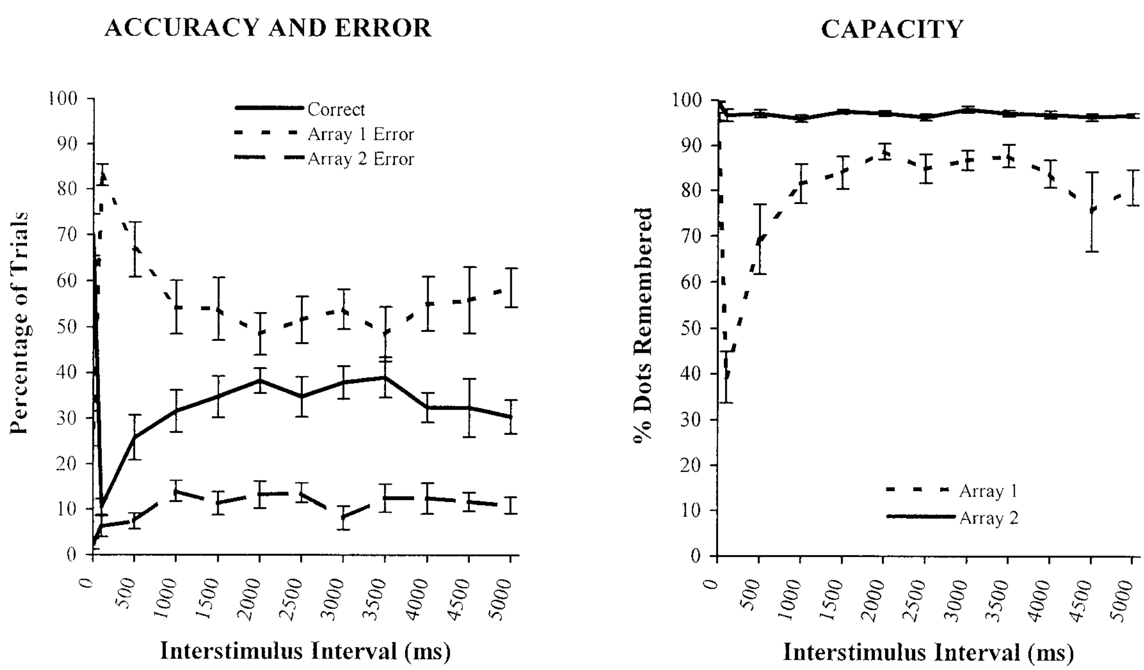

Figure 4. Results of Experiment 4. Left: Mean correct responses, Array 1 errors, and Array 2 errors (with standard error bars) as the percentage of trials on which they occur as a function of interstimulus interval (ISI). Right: Percentage of dots from each array (with standard error bars) that was remembered as a function of ISI.

cost of increasing the ISI from 0 to 100 ms was the same in both experiments $t(7) = -0.2$, $p = .85$. The trends in performance after 100 ms were highly similar in both experiments, differing only in magnitude. That is, accuracy increased in a curvilinear fashion when the ISI exceeded 0 ms, indicating that image–percept integration occurred for 12-dot patterns as well as for 7- and 8-dot patterns. However, a 2 (experiment) × 12 (ISI) repeated measures ANOVA that assessed potential differences in accuracy between Experiments 1 and 4 revealed a reliable main effect of experiment, $F(1, 7) = 47.6$, $MSE = 7.25$. This indicates that accuracy in Experiment 1 was reliably higher than in Experiment 4. In addition, a reliable Experiment × ISI interaction, $F(10, 70) = 4.9$, $MSE = .702$, demonstrated that in Experiment 4, increases in ISI did not produce increases in accuracy as dramatic as those in Experiment 1. Together these results suggest that integrating two 12-dot patterns was more difficult than integrating 7- and 8-dot patterns.

Errors

As a percentage of total trials, Array 1 errors occurred, on average, on 55% of the trials and accounted for 83% of all errors. Thus, in Experiments 1 and 4, Array 1 error was directly linked to accuracy, explaining more than 85% of the variance in accuracy across ISI. Array 2 errors, as a percentage of total trials, occurred, on average, on only 10% of the trials and accounted for 17% of all errors. As in Experiment 1, accuracy and Array 2 error were independent ($r = .02$).

Capacity

The first array. On average, 9.6 dots (80%) presented in the first array were remembered across the ISI. The capacity of the first array varied as a function of ISI, $F(11, 77) = 14.8$, $MSE = 1.13$. As in Experiments 1 and 3, capacity was greatest when the ISI was 0 ms ($M = 11.5$ dots, or 96%) and lowest when the ISI was 100 ms ($M = 4.1$ dots, or 34%). After 100 ms, a curvilinear increase in capacity was observed. Capacity increased until the ISI reached 1,500 ms, after which it attained a mean level of 10.1 dots (84%).

A 2 (experiment) × 12 (ISI) repeated measures ANOVA explored potential differences in the percentage of dots remembered from the first array in Experiment 1 and Experiment 4. A reliable main effect of experiment was observed, $F(1, 7) = 6.0$, $MSE = 5.50$. However, experiment did not interact with ISI, $F(11, 77) = 1.2$. That is, a lesser percentage of the first array was remembered in Experiment 4 (80%) than in Experiment 1 (88%). However, this decrease must be considered in light of the increased number of dots in the first array in Experiment 4—88% of 7 dots (6.2 dots) is less than 80% of 12 dots (9.6 dots). This result indicates that the positions of more dots were remembered in the 12-dot array than in the 7-dot array. In addition, the positions of more dots were also forgotten in the 12-dot array (2.4 dots) than in the 7-dot array (0.7 dots). Obviously, not all dots in Experiment 4 were remembered, suggesting that the 12-dot array exceeded the capacity limitations of the representation.

The second array. The amount of information retained from the second array was, on average, 11.7 dots (97%). A marginal effect of ISI was observed, $F(11, 77) = 1.8$, $MSE = .041$, $p = .06$. However, no reliable interpretable trends were found.

A 2 (experiment) × 12 (ISI) repeated measures ANOVA explored potential differences in the percentage of dots remembered from the second array in Experiment 1 and Experiment 4. A reliable main effect of experiment was observed, $F(1, 7) = 11.2$, $MSE = .073$. However, experiment did not interact with ISI, $F(11, 77) < 1$. Although a statistically lower percentage of the second array was remembered in Experiment 4 (97%) than in Experiment 1 (98%), capacity for the second array was virtually at ceiling in both experiments, and the difference in the capacity of the second array (expressed in percentage terms) was negligible.

Temporal Characteristics

Spline regression showed that capacity for the first array increased until the ISI reached 1,100 ms, after which capacity was constant. As in Experiment 1, the rather long time course suggests that dissipation of backward masking as ISI increases cannot explain the entire increase in capacity. Again, we interpret this as evidence that a slow process is responsible for the increase in performance, which required approximately 1,100 ms to complete. This time course is consistent with the hypothesis that a representation of the first array is generated and maintained during the ISI and then subsequently integrated with incoming perceptual information.

In Experiment 1, the estimated time course of image generation for the 7-dot array was 1,300 ms. In Experiment 4, however, the estimated time to generate a 12-dot image was 1,100 ms. To address the reliability of this difference, the estimate of the time required for capacity to reach asymptote for each subject was computed using the spline regression procedure. The t tests showed that the difference between the estimates attained in Experiments 1 and 2 were unreliable. Therefore, it cannot be concluded that the time course of image–percept integration differed in the two experiments.

Summary

The trends in accuracy, error, capacity, and time course were consistent with those in Experiment 1. Experiment 4 therefore provides additional support for the image–percept integration hypothesis because, despite memory demands that exceeded the capacity of short-term verbal memory, subjects were able to perform the integration task at long ISIs and their representations of the first array contained the equivalent of 10 of the 12 dots it contained. However, accuracy was lower (and error more common) than in Experiment 1. Additionally, capacity for the first array decreased, in terms of the percentage of dots remembered, when the array size increased from 7 to 12 dots. These results indicate that capacity for the image representation may be limited. On the basis of these experiments, it is not clear whether this limitation is one of absolute number of items that can be encoded or of spatial resolution. Studies of visual short-term memory often show limitations in the number of features that can be remembered, although these estimates often vary with task demands (Irwin, 1992; Irwin & Andrews, 1996; Luck & Vogel, 1997; Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Rensink, 2000; Sperling, 1960; Vogel, Woodman, & Luck, 2001). In addition, imagery studies have suggested that the visual buffer upon which images are generated and maintained has a fixed resolution or grain such that smaller or more densely packed features may lead to a less-clear image (e.g., Kosslyn & Alper, 1977).[9] Future research could examine this issue more closely.

Experiment 5

Experiment 5 had two goals. One was to obtain an estimate of the time required to integrate two 12-dot arrays under conditions set for perceptual integration (0-ms ISI) and image–percept integration (2,000-ms ISI) for purposes of comparison with the estimates obtained in Experiment 2, in which a 4 × 4 grid was used. Of interest was whether the time course of image–percept integration would change if more information had to be integrated. The second goal was to replicate the major findings of Experiment 4 (in terms of accuracy, error, and capacity) using a new set of naive, unpracticed subjects, because the subjects in Experiment 4 had participated in Experiment 1 and thus might have been influenced in some fashion by that experience.

Method

Subjects

Eight members of the University of Illinois community participated after providing informed consent. None of the subjects in Experiment 5 participated in Experiments 1–4. All subjects were naive with respect to the experimental hypotheses and were paid $10 for participating.

Stimuli

Stimuli were the same as in Experiment 4.

Apparatus

The apparatus was the same as in Experiment 1.

Design and Procedure

Except for the change in the construction and dimensions of the grid and dot arrays, the design and procedure were the same as in Experiment 2.

Results and Discussion

In general, measures of performance such as accuracy, error, and capacity closely followed those observed in Experiment 4, which also used a 5 × 5 grid. The time course of integration, however, was the same as that observed in Experiment 2.

Accuracy, Error, and Capacity

No effect of response delay was observed on accuracy, Array 1 error, Array 2 error, or capacity for either ISI. On average, when the ISI was 0 ms, accuracy was 75% ($SEM = 0.7$); Array 1 errors occurred on 22% of trials ($SEM = 0.7$); Array 2 errors occurred on 3% of trials ($SEM = 0.10$); 11.7 dots ($SEM = 0.01$), or 98%, were remembered from the first array; 11.9 dots ($SEM = 0.01$), or 99%, were remembered from the second array. These results replicated those in Experiment 4 where perceptual integration was characterized by 70% accuracy, 29% Array 1 error, 1% Array 2 error, 96% image capacity, and 99% percept capacity.

When the ISI was 2,000 ms, accuracy was 30% ($SEM = 0.9$); Array 1 errors occurred on 60% of trials ($SEM = 1.0$); Array 2 errors occurred on 10% of trials ($SEM = 0.30$); 10.0 dots ($SEM = 0.20$), or 83%, were remembered from the first array; 11.7 dots ($SEM = 0.07$), or 97%, were remembered from the second array. These results also replicated those in Experiment 4, where aymptotic performance was characterized by 35% accuracy, 53% Array 1 error, 12% Array 2 error, 84% image capacity, and 97% percept capacity. In addition, comparisons across ISI indicated reliable differences in accuracy, Array 1 error, Array 2 error, and capacity that mirrored similar comparisons in Experiment 4.

Temporal Characteristics

As in Experiment 2, analyses focused on trials on which a correct response was given. Mean RT by ISI and response delay are illustrated in Figure 5. Separate, one-way ANOVAs explored the effect of response delay on RT within each ISI. Each of these analyses is discussed in turn.

RT when the ISI was equal to 0 ms. When the ISI was 0 ms, RT did vary by response delay, $F(11, 77) = 35.5$, $MSE = 3{,}678$, and single-degree-of-freedom polynomial tests showed a reliable negative-slope linear trend in the RT data, $F(1, 7) = 47.1$, $MSE = 16{,}110$. A reliable quadratic trend, $F(1, 7) = 121.4$, $MSE = 4{,}406$, indicated that the decrease in RT as response delay increased did not remain constant across all response delays. Specifically, linear spline regression indicated that RT decreased at a rate of 78 ms for every 100-ms increase in response delay through 460 ms, after which RT increased at a rate of 5 ms for every 100-ms increase in response delay. The overall fit of this model was very good, as indicated by a multiple R of .84. Thus, RT reached floor at approximately 460 ms.

RT when the ISI was equal to 2,000 ms. When the ISI was 2,000 ms, RT also varied by response delay, $F(11, 77) = 7.3$, $MSE = 37{,}372$. Single-degree-of-freedom polynomial tests showed a reliable negative-slope linear trend in the RT data, $F(1, 7) = 29.6$, $MSE = 83{,}047$, as well as a reliable quadratic trend, $F(1, 7) = 15.7$, $MSE = 15{,}448$, indicating that the decrease in RT as response delay increased did not remain constant across all response delays. Specifically, linear spline regression indicated that RT decreased at a rate of 49 ms for every 100-ms increase in response delay through 820 ms, after which RT increased at a rate of 6 ms for every 100-ms increase in response delay. Although the variability in this condition was higher in Experiment 5 compared with Experiment 2, the overall fit of this model was still good, as

[9] During the review process, we learned of a study that used a procedure similar to ours in Experiment 3 but that obtained different results (Zuvic, Visser, & Di Lollo, 1999). This study also used a dot-matrix task to study temporal integration, and it varied the duration of Array 1 up to 1,200 ms. Zuvic et al. found that accuracy declined as duration increased to 100 ms but did not recover at longer durations. There were several procedural differences between the two studies, however, the most striking of which was the visual angle of the displays (18° in our study vs. 2° in the Zuvic et al., 1999, study). It is possible that the difference in results between the two studies was due to subjects being unable to generate a high-resolution image of Array 1 in the Zuvic et al. study, although other explanations are possible as well. This is a topic for future research.

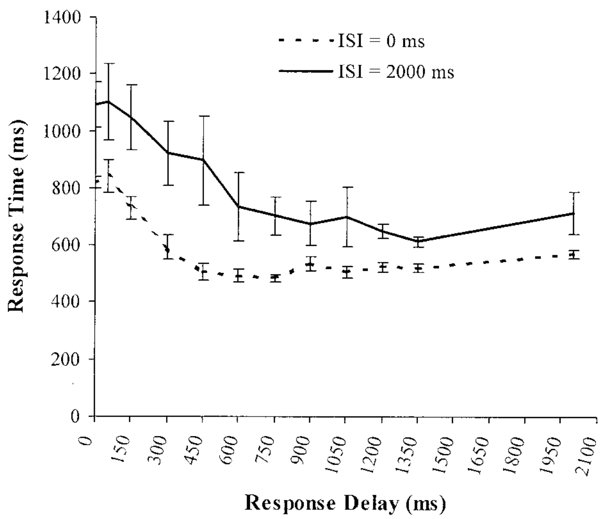

Figure 5. Results of Experiment 5. Mean response times and standard errors by interstimulus interval (ISI) and response delay.

indicated by a multiple *R* of .55. Thus, RT reached floor at approximately 820 ms.

Discussion

As would be expected, accuracy, error, and capacity measures replicated the results of Experiment 4. This indicates that, within these measures, the performance of subjects in Experiment 4 was not significantly influenced by the additional practice of completing Experiment 1 first.

Temporal analyses showed that when the ISI was 0 ms, integration and response selection required approximately 460 ms, whereas when the ISI was 2,000 ms, approximately 820 ms was needed. These estimates are highly similar to those obtained in Experiment 2, where perceptual integration required 500 ms and image–percept integration required 800 ms. Thus, the time required to integrate the elements of a 5×5 array is very similar to the time required to integrate the elements of a 4×4 array, providing support for the hypothesis that images and percepts are integrated directly in some form of immediate analog representation rather than through a time-consuming process of image–image integration or verbal mediation.

In summary, Experiment 5 demonstrated that, despite somewhat lower levels of accuracy using the larger grid, the integration of the image generated from the first array was directly integrated with the perceptual information contained in the second array. The reduction in accuracy, then, is probably due to a limitation in the amount of information that can be retained from the first array, rather than from a less efficient integrative process that combines the information that was retained in the visual buffer.

General Discussion

A large volume of research has suggested that visual imagery and visual perception share a common neural and cognitive architecture. Therefore, one would expect images and percepts to interact in various ways. In this article we obtained evidence that visual images and visual percepts can be integrated to form a new composite representation, and we investigated some of the fundamental properties of this integration process.

Our experiments used a temporal integration paradigm whereby a stimulus is divided into two spatial parts that are presented serially, separated by some time delay. Consistent with previous research investigating perceptual integration, we found that accuracy in combining the two parts of the display was very high when the temporal delay separating the two stimulus parts was 0 ms, but accuracy plummeted when a temporal delay of 100 ms separated the two stimulus parts. However, we found that if the temporal asynchrony of the arrays was extended beyond 100 ms, task performance improved and in some cases approached the levels observed during perceptual integration. Moreover, two temporal properties of performance indicate that this improvement is the result of the direct integration of incoming perceptual information with actively generated and maintained visual images. First, in order for performance to improve to ceiling, more than 1,000 ms was required between array presentations, ruling out explanations based upon conventional masking. This prolonged time course suggests that a slow process was at work that gradually built a more complete representation of the first array. Furthermore, subjects continued to perform quite well even when the limits of nonspatial memory were exceeded. Second, once the representation was generated, the integration and response selection processes together required only an additional 300 ms compared with perceptual integration. This suggests that subjects did not have to form a separate image of the second array, but rather that perceptual information was directly combined with the image already in the visual buffer.

Of critical importance to a complete understanding of image–percept integration is how images and percepts are represented in the visual buffer. In Kosslyn's (1980, 1995, 1997) theory, the visual buffer is analogous to a screen that receives projections from either the eyes (perception) or memory (imagery). If information from each source is simultaneously projected onto the visual buffer, the resulting representation would contain information from each source. Of course, in actuality, representations in the visual buffer must take a very different form. In Kosslyn's view, the visual buffer corresponds to retinotopically mapped areas in the visual cortex within which neural patterns of activity are elicited that correspond to the properties of the seen object. The behavioral and neural similarities of image- and percept-based representations reviewed in the introduction indicate that images and percepts elicit the same patterns of activity in the buffer and are treated equally by processing units further downstream. This similarity in representation, regardless of source, is what would allow the direct integration of incoming percepts and already present images at the level of the visual buffer (for a thorough discussion of the visual buffer, see Kosslyn, 1994).

Given that an image is generated in the visual buffer, what is the nature of the image–percept integration process? That is, how are the image and percept fused? The literature on perceptual integration suggests two possibilities. One class of models focuses on mechanisms that determine retention, or visibility, of the first array (e.g., Eriksen & Collins, 1967; Di Lollo & Hogben, 1985, 1987; Groner et al., 1990). According to this approach, a stimulus remains visible for a short period of time following a brief visual presentation (i.e., visible persistence). Visible persistence degrades rapidly and, left to its own devices, terminates within 100–150 ms (see Coltheart, 1980). In temporal integration experiments where a second percept quickly follows the first, the loss of visible persistence is accelerated because of backward metacontrast masking generated by the second stimulus (Di Lollo & Hogben, 1987; Groner et al., 1990; for a review of early experiments, see Lefton, 1973, or more recently, Enns & Di Lollo, 1997; Lachter & Durgin, 1999). Thus, less information from the first array becomes available and performance decreases sharply.

In contrast, other models emphasize the mechanism or process that integrates the two percepts. For example, Dixon and Di Lollo (1994; see also Di Lollo et al., 1994) proposed that performance in a temporal integration task depends on an integrated representation of the two arrays. The probability (or the degree) that two arrays are integrated depends on the temporal correlation of their visual responses. As the delay between the two percepts increases, the temporal correlation between their evoked responses in the visual system decreases, leading to a drop in performance. In this account, the masking effect is not directly relevant.

Both approaches are consistent with findings in perceptual integration tasks. Similarly, performance in our tasks can also be explained using variations of the two approaches. The information integration approach can easily explain the present findings, if we assume that an image representation of the first array can be integrated with ongoing visual responses of the second array, in basically the same way two visual responses are integrated. For example, as two visual arrays are presented further and further apart in time, the leading array is transformed into an image representation in a visual buffer. Thus, the initial drop of performance can be due to a decreased level of integration between two percepts, whereas the gradual increase reflects a similar but separate process that occurs between a percept and an image.

Indeed, in the experiments reported here, as the duration of the delay between the two arrays increased beyond the temporal limits within which perceptual integration occurs, accuracy improved because the amount of information remembered from the first array increased, leading to fewer errors associated with the leading array. In addition, the time course of the capacity increase was consistent with the speed at which information enters the visual buffer from memory (i.e., *image formation*), which has been shown to be slow—exceeding 1,000 ms (Kosslyn et al., 1985, 1988; Weber & Harnish, 1974). Furthermore, because Experiment 3 showed that increasing the duration of the first array, which also disrupts perceptual integration, enabled subjects to gradually remember more information from the first array, it seems that subjects can form this image either during or after the offset of the stimulus, assuming that such generation is not interrupted by the onset of a subsequent stimulus.

According to an information retention–masking approach, on the other hand, the pattern of performance may be due to masking, which exerts its maximum effect when the delay between the visual percepts is approximately 100 ms and gradually dissipates afterward. However, no known mechanisms of masking have properties that match those in our data. For example, if the improvement in capacity for the first array was primarily due to dissipation of backward masking, the time course of this improvement should be consistent with that of backward masking, which is estimated to be within 300 ms (e.g., Averbach & Coriell, 1961; Breitmeyer, 1980, 1984; Breitmeyer & Ganz, 1976; Kahneman, 1968; Lefton, 1973; Matin, 1975; Turvey, 1973). However, increases in the capacity of the first array did not reach asymptote until the ISI exceeded 1,000 ms, suggesting that the dissipation of masking is not the only contributor to the increase in performance, even if backward masking does occur. Nonetheless, an approach focusing on interruption of information processing of the first array is still possible, although a new type of masking is needed to explain the present findings.

In addition to providing evidence for an image–percept integration process, our data also suggest new insights into the properties of visual images themselves. Consistent with the definitions of Kosslyn (1980, 1995, 1997) and Farah (1984, 1985, 1988), our data suggest that images maintain information about spatial layout in an arraylike format; however, there appears to be a limit on how much information can be maintained in an image. In particular, the results of Experiments 4 and 5 indicate that the positions of approximately 10 dots can be maintained in an image for subsequent integration with incoming perceptual information. Perhaps surprisingly, this estimate of image capacity does not seem to match that of visual short-term memory (VSTM), where one might expect visual images to reside. Several studies have suggested that visual short-term memory seems to have a limit of about 4–5 items (e.g., Irwin, 1992; Irwin & Andrews, 1996; Luck & Vogel, 1997; Phillips, 1974; Pylyshyn & Storm, 1988; Sperling, 1960), yet our subjects managed to remember up to about 10 dots from the first array. Of course, this apparent conflict may not exclude the possibility that images are maintained in VSTM because chunking can occur, which may inflate the estimated capacity in our tasks.

A more intriguing possibility, however, is that memory for location may have different capacity limitations than memory for features such as color, orientation, or identity, which have often been the measures used to determine the capacity of VSTM. One recent contrast of capacity estimates for features and locations is Rensink's (2000) finding using a change detection paradigm that memory for object orientation was approximately 5 items, whereas memory for object location was approximately 9 items, consistent with the 10-item estimate obtained from Experiments 4 and 5. Furthermore, capacity may vary with task demands that may alter the amount of time available for encoding as well as the strategy with which encoding takes place. In the present paradigm, capacity could be relatively high because enough time was given to generate an adequate image of the to-be-remembered information.

One particular property of performance in our experiments, however, seems inconsistent with previous findings in the literature concerning the maintenance and decay of information in VSTM. For example, Phillips (1974) estimated the time course for the decay of information from VSTM by serially presenting subjects with meaningless patterns of black and white squares (i.e., lacking semantic or verbal representations in long-term memory) constructed within a matrix of some size (e.g., 4×4, 6×6, or 8×8), separated by ISIs ranging up to 9,000 ms. The matrices were either identical, or they differed by a single square, and subjects compared the patterns by making same–different judgments. Accuracy on this task decreased as the ISI separating the patterns increased. This pattern of results indicates that VSTM, in addition to being limited, is subject to decay. In all of the experiments reported here, however, performance after 100 ms (measured by either accuracy or capacity) did not decrease once it reached asymptote. Thus, some process enabled subjects to maintain information from the first array, rendering it resistant to decay. It appears that in our task, an image can be formed and actively maintained beyond the limits of VSTM decay. This contrast between the present results and those of Phillips suggest that the characteristics of active working memory may vary depending on the strategies used by subjects while completing different tasks. How different strategies alter the properties of memory is not well understood, however, and constitutes an interesting line of future research.

Another somewhat surprising result is that the time course of image formation did not seem to differ for the 7- and 12-dot arrays. From either a memory-encoding point of view or an image formation point of view, one might expect that it would take longer to process a 12-dot array than a 7-dot array. Indeed, a wide variety of past research has shown that processing time increases as items are added to visual displays or memory sets (e.g., images: Kosslyn, 1980; memory sets: Sternberg, 1966; visual search: Treisman & Gelade, 1980). The fact that it did not take longer in Experiment 4 for performance to reach asymptote suggests that the dots within each array might not be encoded separately but might be chunked by the visual processing system. Furthermore, the same number of chunks may be formed for both the 7- and 12-dot arrays, leading to the similar time course in the two experiments. Future research needs to investigate these issues in more detail to reveal more about the nature of image representations.

Finally, our results suggest a possible resolution to an apparent paradox in the literature on image–percept interactions. As we noted in the introduction, some researchers have found that visual images interfere with visual perception (e.g., Craver-Lemley et al., 1997; Craver-Lemley & Reeves, 1992; Perky, 1910; Reeves, 1981; Segal, 1971), whereas others have found that visual images facilitate visual perception (e.g., Freyd & Finke, 1984; Peterson & Graham, 1974). A potential solution to this apparent conflict arises from our demonstration that image-based and perception-based representations can be integrated in the visual buffer. Given such a process, depending on the nature of the task or the kind of image formed, either interference or facilitation could be observed. For example, should an image that is maintained in the buffer be integrated with incoming perceptual information, the perceptual stimulus will be embedded in the image in the resultant representation, which might impair its detection (e.g., the Perky effect). In such a case, imagery would interfere with visual detection. On the other hand, certain perceptual tasks may benefit from the image if the image provides a useful context. For instance, imaging a square around two lines might aid in determining which line is longer, an effect that would be observed if the square were visually perceived (Freyd & Finke, 1984), because the square can provide a reference frame for the perceptual comparison task. In those cases when the image and the percept are integrated, the resulting representation is potentially richer and more useful, thus making judgments about visual percepts easier.

References

Averbach, E., & Coriell, H. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal, 40,* 309–328.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition, 20,* 449–455.

Breitmeyer, B. G. (1980). Unmasking visual masking: A look at the "why" behind the veil of the "how." *Psychological Review, 87,* 52–69.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* New York: Oxford University Press.

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review, 83,* 1–36.

Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 317–328.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183–228.

Craver-Lemley, C. E., Arterberry, M. E., & Reeves, A. (1997). The effects of imagery on vernier acuity under conditions of induced depth. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 3–13.

Craver-Lemley, C. E., Arterberry, M. E., & Reeves, A. (1999). "Illusory" illusory conjunctions: The conjoining of features of visual and imagined stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1036–1049.

Craver-Lemley, C. E., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review, 99,* 633–649.

Darlington, R. B. (1990). *Regression and linear models.* New York: McGraw-Hill.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109,* 75–97.

Di Lollo, V., & Hogben, J. H. (1985). Suppression of visible persistence. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 304–316.

Di Lollo, V., & Hogben, J. H. (1987). Suppression of visible persistence as a function of spatial separation between inducing stimuli. *Perception & Psychophysics, 41,* 345–354.

Di Lollo, V., Hogben, J. H., & Dixon, P. (1994). Temporal integration and segregation of brief visual stimuli: Patterns of correlation in time. *Perception & Psychophysics, 55,* 373–386.

Di Lollo, V., & Moscovitch, M. (1983). Perceptual interference between spatially separate sequential displays. *Canadian Journal of Psychology, 37,* 414–428.

Dixon, P., & Di Lollo, V. (1994). Beyond visible persistence: An alternative account of temporal integration and segregation in visual processing. *Cognitive Psychology, 26,* 33–63.

Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science, 8,* 135–139.

Eriksen, C. W., & Collins, J. F. (1967). Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology, 74,* 476–484.

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. *Cognition, 18,* 245–272.

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General, 114,* 91–103.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307–317.

Farah, M. J., Soso, M. J., & Dasheiff, R. M. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 241–246.

Finke, R. A. (1980). Levels of equivalence in imagery and perception. *Psychological Review, 87,* 113–132.

Finke. R. A. (1985). Theories relating mental imagery to perception. *Psychological Bulletin, 98,* 236–259.

Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 244–264.

Finke, R. A., & Kurtzman, H. S. (1981a). Area and contrast effects upon perceptual and imagery acuity. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 825–832.

Finke, R. A., & Kurtzman, H. S. (1981b). Mapping the visual field in mental imagery. *Journal of Experimental Psychology: General, 110,* 501–517.

Finke, R. A., & Pinker, S. (1982). Spontaneous imagery scanning in mental exploration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 142–147.

Freyd, J. J., & Finke, R. A. (1984). Facilitation of length discrimination using real and imagined context frames. *American Journal of Psychology, 97,* 323–341.

Goldenberg, G., Podreka, I., Steiner, M., Willmes, K., Seuss, E., & Deecke, L. (1989). Regional cerebral blood flow patterns in visual imagery. *Neuropsychologica, 27,* 641–664.

Groner, R., Groner, M. T., Bischof, W. F., & Di Lollo, V. (1990). On the relation between metacontrast masking and suppression of visible persistence. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 381–390.

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 307–317.

Irwin, D. E., & Andrews, R. V. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 125–155). Cambridge, MA: MIT Press.

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin, 70,* 404–425.

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics, 14,* 90–94.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology, 7,* 341–370.

Kosslyn, S. M. (1976). Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval times. *Memory & Cognition, 4,* 291–297.

Kosslyn, S. M. (1978). Measuring the visual angle of the mind's eye. *Cognitive Psychology, 10,* 356–389.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: MIT Press.

Kosslyn, S. M. (1987). Seeing and imaging in the cerebral hemispheres: A computational approach. *Psychological Review, 94,* 148–175.

Kosslyn, S. M. (1994). Image and brain: The resolution of the imagery debate. Cambridge, MA: MIT Press.

Kosslyn, S. M. (1995). Mental imagery. In S. M. Kosslyn & D. N. Osherson (Eds.), *An invitation to cognitive science: Vol. 2. Visual cognition* (pp. 267–296). Cambridge, MA: MIT Press.

Kosslyn, S. M. (1997). *Image and brain.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M., & Alper, S. N. (1977). On the pictorial properties of visual images: Effects of image size on memory for words. *Canadian Journal of Psychology, 31,* 32–40.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., et al. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience, 5,* 263–287.

Kosslyn, S. M., Ball, T., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of visual scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 47–60.

Kosslyn, S. M., Cave, C. B., Provost, D., & von Gierke, S. (1988). Sequential processes in image generation. *Cognitive Psychology, 20,* 319–343.

Kosslyn, S. M., & Koenig, O. (1992). *Wet mind: The new cognitive neuroscience.* New York: Free Press.

Kosslyn, S. M., Reiser, B. J., Farah, M. J., & Fliegel, S. L. (1985). Generating visual images: Units and relations. *Journal of Experimental Psychology: General, 112,* 278–303.

Lachter, J., & Durgin, F. H. (1999). Metacontrast masking functions: A question of speed? *Journal of Experimental Psychology: Human Perception and Performance, 25,* 936–947.

Lefton, L. A. (1973). Metacontrast: A review. *Perception & Psychophysics, 13,* 161–171.

Loftus, G. R., & Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cognitive Psychology, 35,* 135–199.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Matin, E. (1975). The two-transient (masking) paradigm. *Psychological Review, 82,* 451–461.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology, 8,* 228–246.

Neisser, U. (1976). *Cognition and reality.* San Francisco: Freeman.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition, 3,* 635–647.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology, 21,* 422–452.

Peterson, M. J., & Graham, S. E. (1974). Visual detection and visual imagery. *Journal of Experimental Psychology, 103,* 509–514.

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science, 12,* 287–292.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16,* 283–290.

Pinker, S. (1980). Mental imagery and the third dimension. *Journal of Experimental Psychology: General, 109,* 354–371.

Pinker, S., & Finke, R. A. (1980). Emergent two-dimensional patterns in images rotated in depth. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 244–264.

Pinker, S., & Kosslyn, S. M. (1978). The representation and manipulation of three-dimensional space in mental images. *Journal of Mental Imagery, 2,* 69–84.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monographs, 79*(1, Pt. 2).

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80,* 1–24.

Pylyshyn, Z. W. (1978). *Imagery and artificial intelligence* In C. W. Savage (Ed.), *Minnesota studies in the philosophy of science* (Vol. 9, pp. 19–56). Minneapolis: University of Minnesota Press.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88,* 16–45.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197.

Reeves, A. (1981). Visual imagery lowers sensitivity to hue-varying, but not to luminance-varying, visual stimuli. *Perception & Psychophysics, 29,* 247–250.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7,* 345–376.

Roland, P. E., & Friberg, L. (1985). Location of cortical areas activated by thinking. *Journal of Neurophysiology, 53,* 1219–1243.

Segal, S. J. (1971). *Imagery: Current cognitive approaches.* San Diego, CA: Academic Press.

Segal, S. J., & Fusella, V. (1970). Influences of imagined pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology, 83,* 458–464.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91,* 441–447.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(11, Whole No. 498).

Sternberg, S. (1966, August 5). High-speed scanning in human memory. *Science, 153,* 652–654.

Tippett, L. J. (1992). The generation of visual images: A review of neuropsychological research and theory. *Psychological Bulletin, 112,* 415–432.

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

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review, 80,* 1–52.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114.

Weber, R. J., & Harnish, R. (1974). Visual imagery for words: The Hebb test. *Journal of Experimental Psychology, 102,* 409–414.

Weber, R. J., & Malmstrom, F. V. (1979). Measuring the size of mental images. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 1–12.

Zuvic, S. M., Visser, T. A. W., & Di Lollo, V. (1999, November). *Integration and segregation of brief visual displays: Masking of a pattern by parts of itself.* Poster session presented at the 40th annual meeting of the Psychonomic Society, Los Angeles.

Appendix A

Capacity Calculation

On any given trial, it is certain that subjects remembered some, but not all, of the dots from each array (if they had remembered every dot, accuracy would have been 100%). The number of dots remembered, and forgotten, from each array can be calculated on the basis of the expected probabilities of correct responses, Array 1 errors, and Array 2 errors. Assuming subjects made a response by randomly choosing among the positions in which they did not remember seeing a dot (i.e., positions associated with the dots they forgot plus the empty cell), there are $N + 1$ possible responses, where N is the number of dots forgotten, of which only 1 is correct. Therefore, the probability of getting a correct response, $P(C)$, is

$$P(C) = \frac{1}{(N+1)}. \quad \text{(A1)}$$

When an incorrect response is made, the probability of making an Array 1 error, $P(E)$, is

$$P(E_i) = \frac{i}{(N+1)}, \quad \text{(A2)}$$

where i is the total number of dots forgotten from the first array. Likewise, the probability of making an Array 2 error, $P(E_p)$ is

$$P(E_p) = \frac{p}{(N+1)}, \quad \text{(A3)}$$

where p is the number of dots forgotten from the second array. In order to calculate values of i and p, the corresponding ratio of observed percentages of errors to correct responses is computed. For example, the number of dots forgotten from the first array (i) is found by taking the ratio of observed percentage of Array 1 errors to the percentage of trials on which correct responses were given:

$$\frac{P(E_i)}{P(C)} = \frac{\frac{i}{(N+1)}}{\frac{1}{(N+1)}} = i. \quad \text{(A4)}$$

By the same logic, the number of dots forgotten from the second array is found by computing the ratio of Array 2 errors to correct responses:

$$\frac{P(E_p)}{P(C)} = \frac{\frac{p}{(N+1)}}{\frac{1}{(N+1)}} = p. \quad \text{(A5)}$$

The number of dots remembered from each array, then, is simply the difference between the number of dots presented in each array and the number of dots forgotten from each array.

(*Appendixes continue*)

Appendix B

Spline Regression Procedure

There is a disadvantage to using ANOVA models to estimate the point at which capacity reaches an asymptote in that the estimates are restricted to ISI values that were chosen a priori. The true estimate for the time required for capacity to reach asymptote, however, is likely to fall between the selected ISI values. One way to address this limitation is to use linear spline regression analyses.

Spline regression fits continuous piecewise functions to data (see, e.g., Darlington, 1990). Rather than observe a single linear trend throughout the entire data space, this procedure allows many linear trends to be fit by dividing the data space into segments. Segment boundaries are defined by the joints of the piecewise function. To find appropriate segment boundaries, the x values of the joints are systematically chosen. For each selection, a set of linear functions that minimizes the error in prediction across the whole data space is calculated. The best fitting piecewise function yields the smallest error term. To assess specific hypotheses with spline regression, various functions within the piecewise set can also be defined a priori. For example, between certain joints the function can be defined to be a horizontal line.

Procedure for Experiments 1 and 4

This procedure was used to estimate the point at which capacity reached asymptote. Because spline regression was used to fit linear functions to curvilinear data (the increase in capacity through 1,500 ms produced a reliable quadratic trend in the ANOVA model), capacity values were first normalized and then a polynomial transformation was used to linearize nonlinear data patterns. Specifically, transformation took the form

$$y = z^a , \tag{B1}$$

where z was the normalized capacity score and a was the minimum value required to produce values of y for which no curvilinear trends were observed through 1,500 ms. A two-part piecewise function with the following form was then fit to the data:

$$f(x) = \begin{cases} mx + b & \text{if} \quad x < x' \\ c & \text{if} \quad x > x' \end{cases} , \tag{B2}$$

where x' defines the joint of the piecewise function and c is a constant defined by the value of $f(x')$:

$$c = mx' + b . \tag{B3}$$

Phrased in terms of the experimental design, x' is the point at which capacity no longer differs by ISI. When the ISI is less than x', the spline regression procedure calculates the best fitting regression line for the capacity data; however, for ISIs greater than x', the trend in the capacity data is considered to be a horizontal line. In 50-ms intervals, all values of x' were selected between 500 and 2,000 ms. The overall piecewise function that minimized the mean square error in prediction was accepted, and the value of x' for that function was accepted as the best estimate of the point at which capacity no longer increased by ISI.

Procedure for Experiments 2 and 5

This procedure was used to estimate the point at which RT reached asymptote. Because two linear functions adequately fit the data (i.e., the RT trends both before and after the joint of the calculated piecewise function were linear) no transformations were required on the RT data. In addition, the analyses could not assume that the second linear function had a slope of 0. Thus, a two-part piecewise function with the following form was directly fit to the data:

$$f(x) = \begin{cases} mx + b & \text{if} \quad x < x' \\ mx' + b' & \text{if} \quad x > x' \end{cases} , \tag{B4}$$

where

$$mx' + b = m'x' + b' . \tag{B5}$$

Phrased in terms of the experimental design, x' is the point at which the RT data changes the manner in which it varies by ISI (i.e., the slope of the trend in RT changes). In intervals of 20 ms, all values of x' were selected. The overall piecewise function that minimized the mean square error in prediction was accepted, and the value of x' for that function was interpreted as the point at which the RT–ISI relationship changed.

Received January 3, 2001
Revision received June 26, 2001
Accepted July 12, 2001 ■