VISUAL COGNITION, 2003, *10* (7), 853–873

Integrating visual images and visual percepts across time and space

James R. Brockmole and Ranxiao Frances Wang
University of Illinois, Urbana-Champaign, USA

Recent studies suggest that visual information can be integrated over a relatively long delay (> 1500 ms) to form a more complete representation (image–percept integration). The current studies investigated whether this process can occur between stimuli that differ in their spatial properties. Participants viewed two dot arrays that filled all but one space in a square or rectangular grid when combined, and reported the missing space. The arrays differed either in size or orientation. Performance reached a comparable level as when spatial properties were matched. However, such performance depends on at least two processes. We suggest an early encoding process and a later image formation/spatial attention reallocation process are required. The flexibility of the image–percept integration process suggests a strong mechanism to form more complete or detailed representations over time, even when the retinal size and orientation of the scene may change between successive views.

The visual world contains more information than can be perceived in a single glance; therefore, observers must gradually build up a representation of the world over time. To do this, what has been seen needs to be analysed in conjunction with what is being seen. In recent research, we examined whether the contents of active short term visual memory (i.e., a visual image) can be integrated with newly arriving perceptual information to form a single representation that contains information from each source (Brockmole, Irwin, & Wang, 2003; Brockmole, Wang, & Irwin, 2002).

To explore the possibility of a visual image–percept integration process, Brockmole et al. (2002) used a task in which observers viewed two complementary dot arrays presented sequentially within a square grid. The dot arrays,

Please address all correspondence to: J. R. Brockmole, Dept. of Psychology, Michigan State University, 129 Psychology Research Building, East Lansing, MI 48824, USA.
E-mail: jim@eyelab.msu.edu

Support for this research was provided by a National Science Foundations Graduate Research Fellowship to J. R. Brockmole and a University of Illinois Research Board Grant to R. F. Wang. We thank Derek Larson, David Kaneshiro, and Nicole Stanzi for their help with date collection.

© 2003 Psychology Press Ltd
http://www.tandf.co.uk/journals/pp/13506285.html DOI:10.1080/13506280344000121

if superimposed, filled all but one space in the grid, and observers tried to identify the location of that empty space. To perform this task successfully, observers must commit the first array to memory, retain it for some period of time, and then combine it with the second array. With such displays, near perfect performance is observed if the delay separating the arrays is less than 50 ms as a lingering perceptual trace, or ''iconic memory'', of the lead array overlaps with the presentation of the second array (a phenomenon called perceptual integration). However, such integration is impossible when the delay is extended to just 100 ms (see Di Lollo, 1980; Di Lollo & Hogben, 1987; Di Lollo, Hogben, & Dixon, 1994; Dixon & Di Lollo, 1994; Loftus & Irwin, 1998 for detailed discussion of perceptual integration). In the past, this finding has been taken as evidence that visual integration is impossible once iconic memory has decayed. However, given that, in the absence of iconic memory, successful performance requires both the formation and retention of a longer lasting representation, integration might fail in these short delay situations because of a failure to consolidate the initial representation into a stable and durable format. If true, longer delays between the two arrays might actually *improve* integration performance. Indeed, Brockmole et al. (2002) showed that performance improved steadily as the delay increased from 100 ms to 1300 ms, at which point accuracy asymptoted and did not decay through delays of 5000 ms. Strikingly, accuracy after 1000–1500 ms, in some cases, approached that observed with no delay at all.

Brockmole et al. (2002) argued that the improvement in accuracy over time was facilitated by the formation of a visual image—a short-term visual memory representation that preserves spatial structure—of the first array during the delay, and that the image could be integrated with later perceived visual information. This interpretation was further supported by the subsequent demonstration that, during the interval separating the arrays, selective spatial attention is deployed to the grid locations previously occupied by a dot in the first array (Brockmole et al., in press). Under this *image–percept integration hypothesis*, the integration of visual images and percepts is possible because the cognitive processing of visual images is very similar to the processing of directly perceived stimuli.

There are both structural similarities in visual images and percepts (see Finke, 1980, 1985 for reviews) and in the neurological substrates involved in the processing of images and percepts (see Farah, 1984, 1988; Kosslyn, 1987; and Tippett, 1992 for reviews). For example, both images and percepts encode spatial relationships among objects (Farah, 1985; Finke & Pinker, 1982; Kosslyn, 1973; Kosslyn, Ball, & Reiser, 1978; Pinker, 1980; Pinker & Kosslyn, 1978; Weber & Harnish, 1974), but are limited in spatial resolution as the amount of information contained in an image is constrained by the imagined object's size and distance (Kosslyn, 1975, 1976, 1978; Weber & Malmstrom, 1979), as well as eccentricity of the imagined field of view (Finke & Kosslyn, 1980; Finke & Kurtzman, 1981). Image- and percept-based representations are both processed in the occipital, posterior superior parietal, and posterior inferior temporal lobes. In the primary visual cortex (occipital lobe), the pattern of activation that arises from perception and imagery are similarly affected by the size of the target object with larger objects activating progressively more anterior areas. Finally, studies of abnormal neurological cases have demonstrated that damage to all of these regions can result in vision and imagery impairment.

These commonalties have led some researchers to hypothesize that visual imagery and visual perception are supported by the same cognitive architecture. For example, Kosslyn (e.g., 1994, 1995, 1997; Kosslyn & Koenig, 1992) argues that within the visual system there exists a ''visual buffer'' that maintains activated visual representations, regardless of whether they arose from memory (imagery) or perception. If visual images and percepts are simultaneously maintained in the visual buffer, then they are able to interact and, in the case of image–percept integration, be combined. Thus, given time to generate and maintain visual images, more complete visual representations of the external world can be constructed by integrating currently perceived stimuli into the image representation.

Thus far, studies of image–percept integration have only examined cases where the to-be-integrated stimuli have the same spatial properties. Thus, like perceptual integration, the integration of visual representations based on sensory input (i.e., perception) and those based on visual memory (i.e., imagery) can be a simple fusion of these representations. However, if the properties of the image and percept do not align, the integration process may be challenged as a simple fusion of the information used to derive the image and that given in the subsequent percept would not yield a useful representation. As a result, the effective integration of an image and percept may not be possible.

One potential way to overcome the changes in the spatial properties of to-be-integrated information is the transformation, or modification, of the initial image to match the stimulus properties of the percept. For example, if the image and percept differ in orientation, then the image may be rotated to match the orientation of the upcoming percept. There is ample evidence in the literature that the representation of an initial visual input can be transformed in order to produce an image that deviates from the original stimulus in some way (e.g., shape, size, orientation, colour, texture, etc.). Kosslyn (1994) has divided such transformations into two classes. The first type is a *motion-encoded* transformation. This kind of transformation is one in which the imaged stimulus itself contains some sort of motion. For example, if one views a moving car they may later be able to imagine that car in motion. The second type, and the class which is the focus of the present experiments, is a *motion-added* transformation. This kind of transformation occurs when a static stimulus is viewed and the observer then imagines it ''moving'' or changing in some way.

Several kinds of motion-added transformations have been examined. For example, Bundesen and Larson (1975) investigated imagined transformations of size. They showed subjects pairs of polygons which they were to judge as either being same or different, regardless of their size. Response times increased linearly as the polygons' size difference increased. This suggests that subjects mentally expanded or contracted one of the figures to match the scale of the other. The linear function describing response time indicates that the figures were scaled at a constant rate and that the manipulated figure passed through all intermediate stages between its initial size and the final resultant size. To investigate transformations of shape, Shepard and Feng (1972) constructed two dimensional patterns that could be folded to form a cube. On each pattern, two arrows were depicted; the subjects' task was to indicate whether the arrows would meet if the pattern were folded into the cube. Subjects were able to perform the task, but response times increased linearly with the number of folds that would have to be made to form the cube, suggesting that subjects imagined performing each of the folds to accomplish the task.

The vast majority of studies on motion-added image transformations, however, have examined mental rotation. In the seminal paper on the topic, Shepard and Metzler (1971) showed subjects pairs of three dimensional figures that were either perfectly aligned or misaligned by three-dimensional rotations in 20° increments. The subjects' task was to indicate whether the figures in each pair were the same. Response times increased linearly with the angle of deviation between the figures. This result suggests that subjects made their judgements by rotating one of the figures in the pair so that its orientation matched that of its partner. This effect has been replicated using letters and numbers (Cooper & Shepard, 1973), hands (Cooper & Shepard, 1975), dot patterns (Corballis & Roldan, 1975), and multiple-object configurations (Hintzman, O'Dell, & Arndt, 1981).

The purpose of the experiments reported here was to explore the possibility of image–percept integration across changes in the spatial aspects of the to-be-integrated stimuli over time—when the image needs to be transformed in order to be meaningfully integrated with a subsequent percept that is presented in a different size scale or orientation. The major questions of interest were whether image–percept integration is possible when transformations are required, and what temporal conditions are important to successful integration in those cases. If the first array is maintained in the visual buffer as an image, and if motion-added transformations such as size scaling and rotation can be applied to the representation, then, given sufficient time, integration of discrete visual stimuli should be possible despite differences in the spatial properties of the to-be-integrated stimuli.

EXPERIMENT 1

The purpose of Experiment 1 was to investigate image–percept integration under cases where the first and second arrays were presented at different size scales. Size scaling poses a challenge to the image–percept integration system. When the first and second arrays are presented in different scales, a strictly veridical representation of array 1 cannot be integrated with array 2. Rather, it must be mentally expanded or contracted to match the size of array 2.

Like Brockmole et al. (2002, 2003), observers were presented with two sequentially presented dot arrays within enclosed square grids separated by a variable interstimulus interval (ISI) and were asked to indicate which grid space was not filled with a dot. During the ISI separating the arrays, however, the grid changed size. The transformation of the grid divided the ISI into two pieces: The time that elasped prior to the transformation and the time that elasped after the transformation. Each of these two components of the ISI may be critical to image–percept integration. First, the time that elapses between the offset of the first array and the enlargement of the grid may be important for extracting information from the initial array required for image generation. If the grid changes size too quickly, such that the first array cannot be adequately encoded, then little information may be retained from which to generate an image. Second, the time that elapses between the enlargement of the grid and the onset of the second array may be important for processing the information committed to memory along the new spatial dimensions. Each of these time components, which will be referred to as early components and late components, respectively, was independently manipulated in Experiments 1A and 1B, respectively, to contrast the contributions of the time available for processing prior to grid scaling and the time afforded for processing following grid scaling and to determine how much time was necessary before and after grid scaling to produce optimal performance.

Experiment 1A

Experiment 1A examined performance as a function of the time available prior to grid scaling. This was accomplished by varying the time that elapsed between the offset of the first array and the presentation of the second, spatially enlarged grid. This will be referred to as the *prescaling* delay. In order to examine performance as a function of the prescaling delay, this condition severely limited the amount of time available for postscaling processing by holding it constant at 100 ms. That is, after the second grid was presented, it was filled with a dot array 100 ms later. Unless postscaling processes require less than 100 ms, this time course does not allow such processes to be completed prior to the onset of the second array. As a result, performance in this condition should largely reflect the time course of prescaling processing and should then reveal how well the arrays can be integrated given only those processes at work.

Method

Subjects. Twelve undergraduates at the University of Illinois participated after providing informed consent. All subjects were naive with respect to the experimental hypotheses and were paid $5 for participating.

Stimuli. Stimuli consisted of two unique dot arrays sequentially presented within two enclosed 4 × 4 square grids, respectively. The centre of the first and second grids coincided with the same point in space. However, the first grid and the dots presented within it were three times smaller than the second grid and array. The first array contained seven dots and the second array contained eight dots. If the two grids were superimposed in space and displayed at the same scale, the arrays would fill all but one square in 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 horizontally (approx. 37° of visual angle) and 27.9 cm (29°) vertically. The first grid subtended 4.4 cm (5°) horizontally and vertically. Each square within the grid subtended 1.1 cm (1.3°), and each dot presented in the arrays subtended 0.9 cm (1°). The second grid subtended 13.2 cm (15°) horizontally and vertically. Each square within the grid subtended 3.3 cm (3.9°), and each dot presented in the arrays subtended 2.7 cm (3°). The display background was light grey, 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 they thought had been unfilled.

Design and procedure. The procedure is shown in Figure 1. Each trial consisted of two sequentially presented dot arrays within enclosed square grids separated by a variable interstimulus interval (ISI). During the ISI separating the arrays, the grid changed size. The first and second arrays together filled all but one position in the grid. Under the constraint that no members of the two arrays occupy the same position in the grid, the dot patterns were randomly generated on each trial. Thus, the location of the missing dot was equally likely to occur in each grid position. Subjects were instructed to identify the position of the missing dot.

Prior to the onset of each trial, an empty 4 × 4 square grid was presented on the left side of the display. When ready to begin the trial, sub-

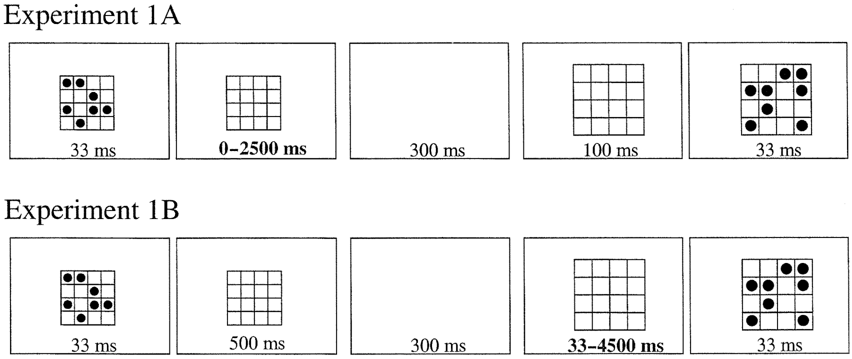

Figure 1. Schematic illustration of the procedure used in Experiment 1.

jects pressed the spacebar. Immediately, the first array of seven dots was presented within the grid for 33 ms. Following array 1 offset, the grid retained its initial size (the prescaling delay) for either 0, 33, 67, 100, 300, 500, 1000, 1500, or 2500 ms whereupon it was erased. Following grid erasure the display remained blank for 300 ms. The (empty) second grid then appeared for 100 ms. The second dot array was then presented for 33 ms. Thus, the ISI separating the arrays was 400, 433, 467, 500, 700, 900, 1400, 1900, or 2900 ms. The 300 ms blank interval inserted between the offset of the first grid and onset of the second grid was done to avoid any potential backward, lateral, or object-substitution masking effects, which are known to dissipate within 300 ms (Averbach & Coriell, 1961; Breitmeyer, 1980, 1984; Breitmeyer & Ganz, 1976; Enns & Di Lollo, 1997; Kahneman, 1968; Lefton, 1973; Matin, 1975; Turvey, 1973).

Trials were divided into blocks based on the prescaling delay. Subjects were informed of the delay prior to the start of each block. Subjects completed 270 trials, divided into blocks of 10. Each delay (nine) occurred equally often. For each subject, the blocks of trials were presented in a different random order. Prior to beginning the experimental trials, subjects completed two practice sessions to familiarize themselves with the experimental procedure. In the first session, subjects completed 40 trials where the first and second grids were both presented at the same size scale. This session consisted of 20 trials each at ISIs of 0 and 2000 ms. In the second session, subjects completed 40 trials identical to the experimental trials. This session consisted of 10 trials at pre-scaling delays of 100, 500, 1000, and 2000 ms each. During the practice sessions, subjects were given feedback concerning the accuracy of their responses. However, during the experimental trials no feedback was given.

Results

Analyses focused on trends in accuracy and error as a function of the prescaling delay. Trends in accuracy and error are illustrated in Figure 2. For all analyses, an alpha level of .05 was adopted as the criterion for statistical reliability.

A one-way repeated measures analysis of variance with the prescaling delay as the independent factor demonstrated that accuracy varied as a function of this delay, $F(8, 88) = 2.94$, $MSE = 96.7$. Single degree of freedom polynomial tests indicated a reliable positive-slope linear trend, $F(1, 11) = 16.9$, $MSE = 116.7$ in the accuracy data, indicating the accuracy improved as the prescaling delay increased.

In terms of interpreting the variation in accuracy across time, one goal was to determine the point at which accuracy no longer increased as a function of the prescaling delay. The ''breaking point'' at which the trend in accuracy changed was assessed using linear spline regression, a methodology that also allows for the interpolation of this breaking point between the predisplacement delay values that were selected a priori. Based on Brockmole et al. (2002), two-part linear functions were fit to the accuracy data by varying the slope of the segments and the joint between the lines. The joint of the best fitting set of functions was accepted as an estimate of the point at which the trend in accuracy changed. Accuracy improved until the prescaling delay was 400 ms (800 ms ISI) over which time accuracy improved from 29% to 36%. The slope of the second segment did not reliably differ from 0, indicating that accuracy reached asymptote and did not differ as a function of the prescaling delay, averaging 39%.

In addition to accuracy, error rates were analysed. Errors were divided into two types. Array 1 errors are those where the errantly selected grid position was previously occupied by a dot from the first array. Array 2 errors, on the other hand, are those where the errantly selected grid position was previously occupied by a dot from the second array. Each error type is a relative indicator of how much information was remembered from each array as better memory should translate to fewer errors. The source of errors was assessed with an emphasis on the strength of the relationship between different types of error and accuracy. As a percentage of total trials, array 1 errors occurred, on average, on 59% of the trials and accounted for 90% of all errors. The Pearson correlation between array 1 errors and correct responses across ISIs was $-.89$, revealing a very high level of dependence between accuracy and array 1 error. This means that 79% of the variance in accuracy is accounted for by array 1 errors and, thus, the improvement in accuracy was mainly due to the formation of a more complete memory representation of the first array. Array 2 errors, however, as a percentage of total trials occurred, on average, on 7% of the trials and accounted for 10% of all errors. The Pearson correlation between array 2 errors and correct

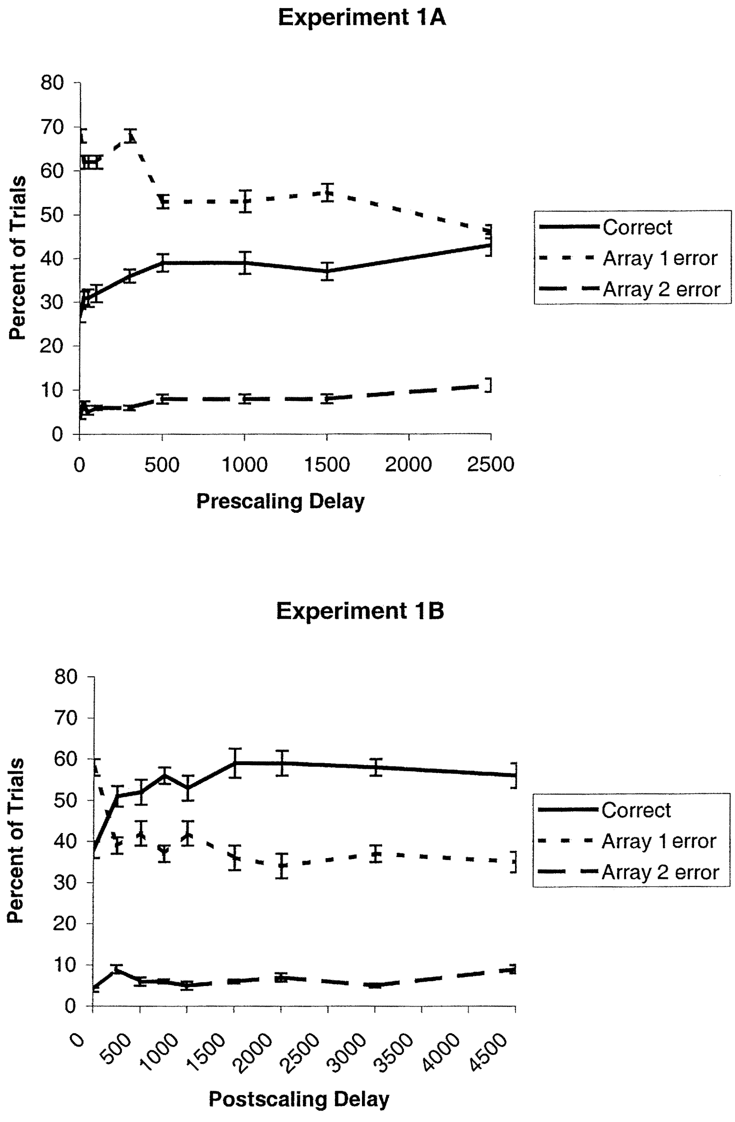

Figure 2. Results of Experiment 1. Correct responses, array 1 errors, and array 2 errors are illustrated as a percentage of trials on which they occurred by pre- and postscaling delay.

responses across ISIs was −.39, revealing lesser dependence between accuracy and array 2 error. Array 2 error accounted for only 15% of the variance in accuracy.

Experiment 1B

Experiment 1B examined performance as a function of the time available for processing after grid enlargement. This was accomplished by varying the time that elapsed between the enlargement of the grid and the presentation of the second array. This will be referred to as the *postscaling delay*.

Method

Subjects. Twelve undergraduates at the University of Illinois participated after providing informed consent. All subjects were naive with respect to the experimental hypotheses and were paid $5 for participating. None of the subjects participated in any other experiment.

Stimuli and apparatus. The same stimuli and apparatus was used as in Experiment 1A.

Design and procedure. The design and procedure were the same as in Experiment 1A except for the following changes (see Figure 1). In Experiment 1B, it was essential to allow early processing to be optimally completed prior to grid enlargement. If prescaling processing is not complete, either it will be continued postscaling, or it will not be completed at all, which may reduce general performance. Thus, if prescaling processing is complete, trends in performance as a function of postscaling time should largely reflect the contribution of postscaling processing on image–percept integration. Ideally, then, the prescaling delay used in Experiment 1B should be the shortest duration possible that allows all prescaling processes to be completed. Based on the results of Experiment 1A, this duration is 400 ms (the time after which performance in the early component condition reached asymptote). However, because different subjects completed the late component condition and because there is certainly individual variance around this 400 ms estimate, the prescaling time in the late component condition was increased to 500 ms to be certain that enough time was provided for prescaling processing.

Thus, following array 1 offset, the grid remained present for 500 ms whereupon it was erased. As in Experiment 1A, this was followed by a 300 ms interval during which the display was blank. The second grid then appeared following a variable delay of either 33, 267, 500, 767, 1000, 1500, 2000, 3000, or 4500 ms (the postscaling delay). Finally, the second dot array was presented for 33 ms. Thus, the ISI separating the arrays was 833, 1067, 1300, 1567, 1800, 2300, 2800, 3800, or 5300 ms. During practice, postscaling delays of 100, 500, 1000, and 2000 ms were used.

Results

Analyses focused on trends in accuracy and error as a function of the post-scaling delay. Trends in accuracy and error are illustrated in Figure 2. A one-way repeated measures analysis of variance with the postscaling delay as the independent factor demonstrated that accuracy varied as a function of this delay, $F(8, 88) = 4.04$, $MSE = 125.6$. Single degree of freedom polynomial tests indicated a reliable positive-slope linear trend, $F(1, 11) = 11.1$, $MSE = 209.9$, in the accuracy data, indicating the accuracy improved as the postscaling delay increased. A linear spline regression was performed on these data as well. This analysis indicated that accuracy improved until the postscaling delay was 350 ms (1150 ms ISI) over which time accuracy improved from 38% to 53%. The slope of the second segment did not reliably differ from 0, indicating that accuracy did not differ as a function of the postscaling delay, averaging 56%.

As a percentage of total trials, array 1 errors occurred, on average, on 40% of the trials and accounted for 85% of all errors. The Pearson correlation between array 1 errors and correct responses across ISIs was $-.95$, revealing a very high level of dependence between accuracy and array 1 error. Approximately 90% of the variance in accuracy was accounted for by array 1 errors, meaning that the improvement in accuracy was mainly due to the formation of a more complete memory representation of the first array. Array 2 errors, however, as a percentage of total trials occurred, on average, on 6% of the trials and accounted for 15% of all errors. The Pearson correlation between array 2 errors and correct responses across ISIs was $-.24$, revealing a lesser degree of dependence between accuracy and array 2 error. Array 2 errors accounted for only 6% of the variance in accuracy.

Performance in Experiment 1B was compared to that in Experiment 1A by collapsing across pre- and postscaling delays that exceeded the point at which accuracy no longer varied (i.e., reached asymptote) and performing a two-sample *t*-test. Because the number of observations in each group was not equal, variances were pooled. This analysis showed that asymptotic accuracy in the late component condition was reliably higher than that in the early component condition, $t(118) = 5.18$. The magnitude of this advantage was, on average, 17%. With respect to error rates, during this time frame, array 1 errors were reliably lower in the late component condition, $t(118) = 4.89$, although array 2 errors were only marginally lower in the late component condition, $t(118) = 1.79$, $p = .08$. The magnitude of these differences were, on average, 15% and 2%, respectively.

Discussion

The overall data from Experiment 1 suggest that image–percept integration is possible when the to-be-imaged stimulus is not presented at the same spatial scale as the subsequent to-be-integrated stimulus. In order for such integration to be possible, subjects had to mentally scale the first stimulus until it matched the spatial dimensions of the second stimulus. As in previous demonstrations of image–percept integration (Brockmole et al., 2002, 2003) timing was critical. The ISI separating the arrays, although an essential factor in allowing image–percept integration to occur, was not the only temporal determinant of performance. The point during the ISI at which the grid changed scale was also a major factor in determining performance. Optimal performance was achieved when the time prior to grid erasure was approximately 400 ms and when the time after scaling was approximately 350 ms. If the pre- or postscaling delays were less than these values, integration was suboptimal. Considering the additional 300 ms blank interval, these values produced an ISI of 1050 ms.

EXPERIMENT 2

Experiment 2 investigated image–percept integration under cases where the first and second arrays are presented at different orientations. When the grids are presented at different orientations, the grid geometry changes. For example, if a horizontally positioned rectangular 5×3 grid is rotated 90° clockwise, what was the upper left space in the first grid is now the upper right space in the second grid. The general logic and methodology was the same as in Experiment 1. In Experiment 2, the grid dimensions were changed so that it would be rectangular. The first grid was oriented horizontally, and the second grid was oriented vertically. Subjects were instructed to imagine that the grid rotated 90° clockwise.

Experiment 2A

Experiment 2A examined performance as a function of the time available prior to grid rotation by varying the time that elapsed between the offset of the first array and the presentation of the second, rotated grid. This is referred to as the prerotation delay.

Method

Subjects. Thirteen undergraduates at the University of Illinois participated after providing informed consent. All subjects were naive with respect to the experimental hypotheses and were paid $5 for participating. None of the subjects participated in any other experiment.

Stimuli. As in Experiment 1, the centre of the first and second grids coincided with the same point in space. The first grid was rectangular in form and presented so that the major axis was horizontal in the image plane; this grid measured 5 squares $\times$ 3 squares and subtended 14.0 cm horizontally (approximately 15.9°) and 8.4 cm (9.6°) vertically. The second grid was also rectangular but it was presented so that the major axis was vertical in the image place; this grid therefore measured 3 squares $\times$ 5 squares and subtended 8.4 cm horizontally (9.6°) and 14.0 cm (15.9°) vertically. For both grids, each square subtended 2.8 cm (3.2°), and each dot presented in the arrays subtended 2.4 cm (2.7°). The arrays each contained seven dots.

Apparatus. The apparatus was the same as in Experiment 1.

Design and procedure. The design and procedure was the same as in Experiment 1A, except that the grid changed orientation between array presentations instead of size (see Figure 3). This change in orientation was viewed by the subject as a rotation of the first grid from horizontal to vertical. Subjects were instructed that the direction of this rotation was 90° clockwise.

Results

Analyses mirrored those performed in Experiment 1A. Trends in accuracy and error conditions are illustrated in Figure 4. A one-way repeated measures analysis of variance with the prerotation delay as the independent factor demonstrated that accuracy varied as a function of this delay, $F(8, 96) = 4.91$, $MSE = 70.6$. Single degree of freedom polynomial tests indicated a reliable positive-slope linear trend, $F(1, 12) = 13.6$, $MSE = 161.0$ in the accuracy data, indicating the accuracy improved as the pre-rotation delay increased. Spline regression analysis indicated that accuracy improved until the prerotation delay

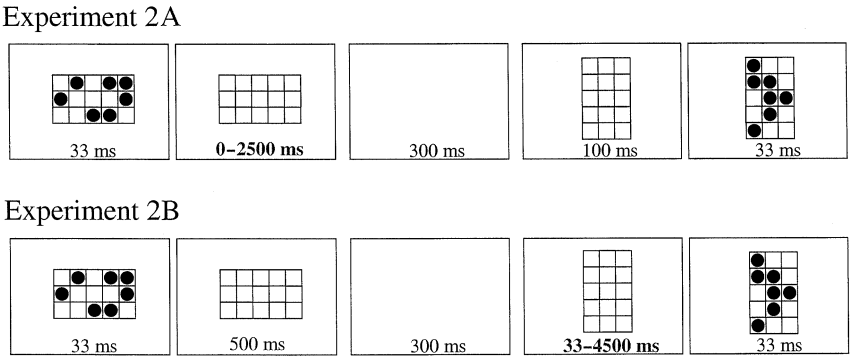

Figure 3. Schematic illustration of the procedure used in Experiment 2.

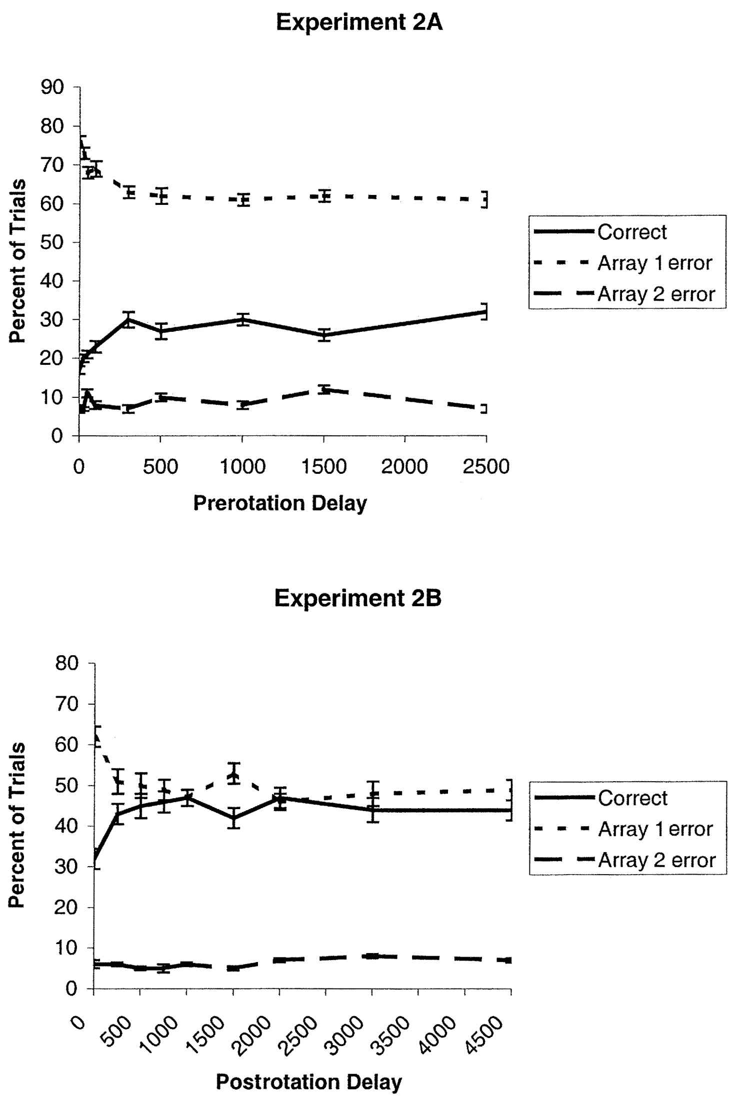

Figure 4. Results of Experiment 2. Correct responses, array 1 errors, and array 2 errors are illustrated as a percentage of trials on which they occurred by pre- and postrotation delay.

was 250 ms (650 ms ISI) over which time accuracy improved from 18% to 30%. The slope of the second segment did not reliably differ from 0, indicating that accuracy did not differ as a function of the prerotation delay, averaging 30%.

As a percentage of total trials, array 1 errors occurred, on average, on 62% of the trials and accounted for 88% of all errors. The Pearson correlation between array 1 errors and correct responses across ISIs was −.89, revealing a very high level of dependence between accuracy and array 1 error. Array 1 error accounted for 79% of the variance in accuracy. This means that the improvement in accuracy was mainly due to the formation of a more complete memory representation of the first array. Array 2 errors, however, as a percentage of total trials occurred, on average, on 10% of the trials and accounted for 12% of all errors. The Pearson correlation between array 2 errors and correct responses across ISIs was −.33, revealing a lesser degree of dependence between accuracy and array 2 error. Array 2 error accounted for only 11% of the variance in accuracy.

Experiment 2B

Experiment 2B examined performance as a function of the time available for processing after grid rotation by varying the time that elapsed between the rotation of the grid and the presentation of the second array. This will be referred to as the *postrotation delay*.

Method

Subjects. Thirteen undergraduates at the University of Illinois participated after providing informed consent. All subjects were naive with respect to the experimental hypotheses and were paid $5 for participating. None of the subjects participated in any other experiment.

Stimuli and apparatus. The stimuli and apparatus was the same as in Experiment 2A.

Design and procedure. The design and procedure were the same as in Experiment 1B except for the type of change (rotation instead of scale change, see Figure 3). Although the results of Experiment 2A indicated that the shortest duration possible that still allowed all prerotation processes to be completed was 250 ms, to account for individual variance around this estimate, the prerotation delay was increased to 500 ms.[1]

Results

Analyses focused on trends in accuracy and error as a function of the postrotation delay. Trends in accuracy and error are illustrated in Figure 4. A one-way repeated measures analysis of variance with the postrotation delay as the independent factor demonstrated that accuracy varied as a function of this delay,

[1] The minimum pretransformation delay that yielded maximum performance in Experiment 2A was more variable than that in Experiment 1A, thus the prerotation delay was increased to a greater degree in this experiment.

$F(8, 96) = 2.64$, $MSE = 97.8$. Single degree of freedom polynomial tests indicated a reliable positive-slope linear trend, $F(1, 12) = 6.77$, $MSE = 86.7$, in the accuracy data, indicating the accuracy improved as the postrotation delay increased. The spline regression indicated that accuracy improved until the postrotation delay was 300 ms (1100 ms ISI) over which time accuracy improved from 32% to 45%. The slope of the second segment did not reliably differ from 0, indicating that accuracy did not differ as a function of the post-rotation delay, averaging 45%.

As a percentage of total trials, array 1 errors occurred, on average, on 50% of the trials and accounted for 89% of all errors. The Pearson correlation between array 1 errors and correct responses across ISIs was −.96, revealing a very high level of dependence between accuracy and array 1 error. Array 1 error accounted for 92% of the variance in accuracy. This means that the improvement in accuracy was mainly due to the formation of a more complete memory representation of the first array. Array 2 errors, however, as a percentage of total trials occurred, on average, on 6% of the trials and accounted for 11% of all errors. The Pearson correlation between array 2 errors and correct responses across ISIs was −.20, revealing a lesser degree of dependence between accuracy and array 2 error. Array 2 errors accounted for only 4% of the variance in accuracy.

Performance in Experiment 2B was compared with that in Experiment 2A by collapsing across pre- and postrotation delays that exceeded the point at which accuracy no longer varied (i.e., reached asymptote) and performing a two-sample *t*-test. Because the number of observations in each group was not equal, variances were pooled. This analysis showed that asymptotic accuracy in the late component condition was reliably higher than that in the early component condition, $t(141) = 5.88$. The magnitude of this advantage was, on average, 16%. With respect to error rates, during this time frame, image errors were reliably lower in the late component condition, $t(141) = 4.92$, as were percept errors, $t(141) = 2.55$. The magnitude of these differences were, on average, 13% and 3%, respectively.

Discussion

The overall data from Experiment 2 suggest that image–percept integration is also possible when the to-be-imaged stimulus is not presented in the same orientation as the subsequent to-be-integrated stimulus. That is, subjects could mentally rotate the first stimulus until it matched the spatial dimensions of the second stimulus prior to integrating the two arrays. Again, two time courses are important. Optimal performance was achieved when the time prior to grid rotation was approximately 250 ms and when the time after rotation was approximately 300 ms. If the pre- or postrotation delays were less than these values, integration was suboptimal. Considering the additional 300 ms blank interval, these values produced an ISI of 850 ms.

GENERAL DISCUSSION

In previous demonstrations of image–percept integration where the first and second arrays were presented within the same grid, observers generated an image of the lead array by allocating visual selective attention and memory to locations that were occupied by dots in the lead array over a period of approximately 1000–1500 ms (Brockmole et al., 2002, 2003). The purpose of the present study was to examine whether image–percept integration is possible when the to-be-integrated visual stimuli differ along various spatial dimensions such as scale and orientation. In order for meaningful integration to occur under these circumstances, the representation of the first stimulus had to be mentally transformed to match the spatial properties of the second stimulus.

Past studies on motion-added transformations of visual images have indicated that images are spatially flexible to the extent that they can be mentally scaled and rotated to produce patterns that were not directly perceived. Thus, similar transformations should be available to manipulate the visual image that maintains the spatial layout of to-be-remembered stimuli. These transformations would allow spatially mismatching stimuli to be integrated, provided that sufficient time is afforded for such transformations to be completed. Indeed, the present study demonstrated that image–percept integration is possible when the spatial properties of the to-be-integrated stimuli do not match. That is, subjects could maintain and manipulate image-based representations in order to enable integration of spatially mismatching visual stimuli. This conclusion has several implications for conceptions of image–percept integration as a means of combining discrete pieces of visual information over time in dynamic displays.

First, in terms of processing time, image–percept integration across scale and orientation is similar to cases where no transformations are required. To produce optimal performance, for scale change, a total ISI of 1050 ms was required between array presentations, including pre- and postscaling delays of 400 ms and 350 ms, respectively. For orientation change, optimal performance required a total ISI of 850 ms, including pre- and postrotation delays of 250 ms and 300 ms, respectively. The time course for image–percept integration when no transformations are required has been observed to range from 1000 to 1500 ms (Brockmole et al., 2002, in press). This suggests that the requirement to transform the image to match the spatial properties of the percept entails no temporal cost.

Second, in terms of accuracy, image–percept integration across scale and orientation is very efficient. Asymptotic performance when the to-be-integrated stimuli were presented in different scales was 56% and at different orientations it was 45% (compare this to a mean asymptotic accuracy level of 63% for 4 × 4 grids where arrays 1 and 2 contained seven and eight dots, respectively (Brockmole et al., 2002). This is impressive considering that chance performance, even assuming perfect memory for the second array, is 12.5%. Thus, accuracy was at least four times greater than that level. In fact, assuming that subjects randomly choose among grid positions that were occupied by forgotten dots and the space that was never filled with a dot, accuracy of 50% means that subjects remembered the positions of all but one dot from the combined arrays.

Third, considering the processing time and performance together, an apparent paradox emerges: Compared to scale transformations, rotation transformations seem to require less time overall (850 ms ISI vs. 1050 ms), and less pre-transformation time (250 ms vs. 400 ms), but led to worse performance. Although consistent with a speed–accuracy tradeoff account, a more intriguing and counterintuitive possibility exists: The more difficult image generation is, the less processing time is required. If the maintenance of information in the image is more difficult during mental rotation than scaling, then more information about the initial stimulus would be lost during rotation. As a result, the generation of the integrable image, which contains less information, would require less time. This explanation can also account for why optimal integration performance in the present experiments is relatively fast compared to prior work examining integration when no transformations are required (1000–1500 ms). Future work should systematically explore this potential tradeoff between the difficulty and time course of image generation within the image–percept integration paradigm.

Fourth, although the interstimulus interval between to-be-integrated stimuli was an necessary factor in image–percept integration, it was not sufficient. In both experiments, optimal integration was dependent on the time before and after the change in scale or orientation took place. That is, given equal ISIs, the distribution of pre- and postchange processing time has a profound impact on performance (see Figure 5). This suggests that various cognitive processes must be permitted to be completed at different points during the ISI separating the arrays. What cognitive processes are occurring prior to and following the spatial changes examined here? We speculate that prior to the change, the to-be-imaged array had to be encoded into a short-term spatial memory store from which an image could be generated. If insufficient time was provided to complete this process, the quality of the image, and therefore integration performance suffered. Following the change, an image with the proper spatial properties needs to be generated. That is, attention would have to be transferred to the new grid in preparation for stimulus processing. If the second array was presented before such allocation was complete, integration also suffered. Future research is required to fully evaluate these hypotheses.

Fifth, considering the current experiments in the light of past work on image–percept integration adds to our understanding of the nature of the image representations used to combine visual stimuli over time. Using a similar dot-integration paradigm, Brockmole et al. (2002) demonstrated that once the image is formed, subsequent perceptual information can be directly fused with it without undergoing a similar image generation process. Then, Brockmole et al.

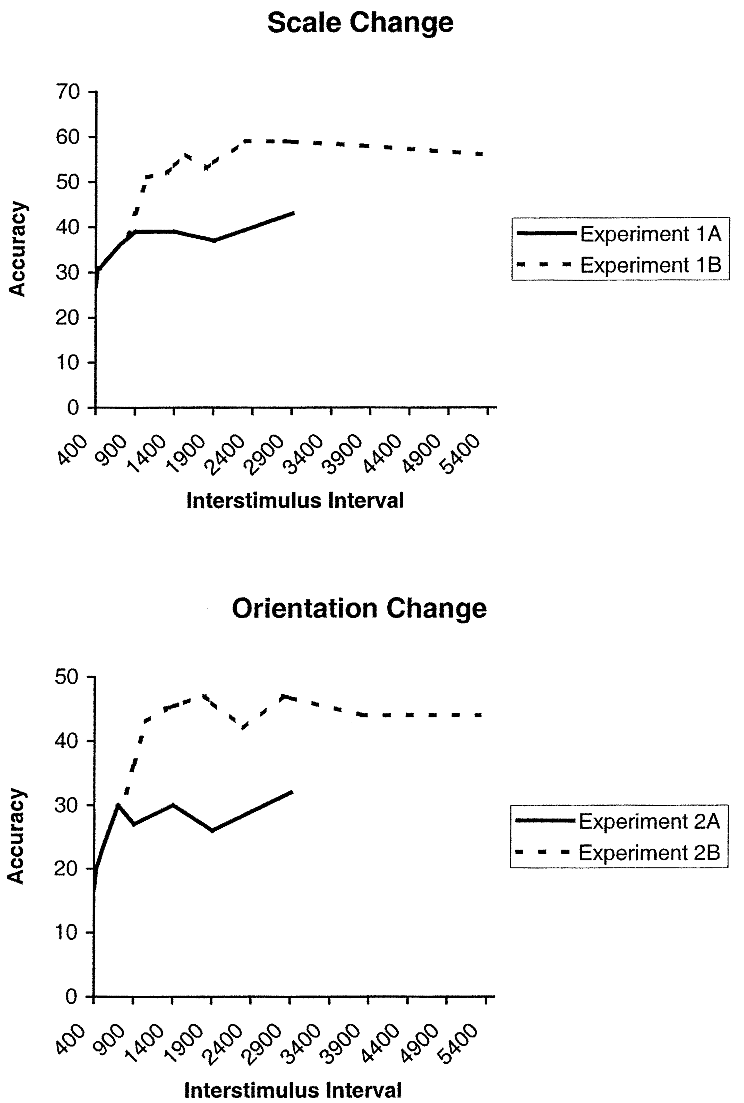

Figure 5. A comparison of accuracy in the early and late component condition for scale and orientation transformations. Accuracy is plotted as a function of the interstimulus interval between the two to-be-integrated arrays.

(2003) showed that that image is created by deploying visual selective attention to those aspects of the grid that were occupied by a dot from the first array. The present experiments suggest that this assignment of attention is not dedicated to particular areas of space independently of the stimulus that occupied it: The image can be modified to account for changes in stimulus structure and ‘‘projected’’ to completely different regions of the display. What remains an interesting question, however, is what kind of information about the stimulus is maintained in the image. For example, the image may encode the individual dots in the first array, treating each of these as a separate element, or it may encode the overall shape or gestalt created by the dots. We are currently using other measures of performance such as eye movements as well and manipulations of stimulus properties such as the homogeneity of the elements or the ease with which the elements can be grouped to examine these possibilities.

Finally, these experiments support the hypothesis that image–percept integration can be used to combine visual information in dynamic scenes. Objects in the visual world can move and change the spatial properties of their images on the retina. In order to combine information obtained at discrete time points while viewing a visual scene, one must be able to deal with changes in the spatial properties of the objects in the scene. As an object becomes larger or smaller as it approaches or retreats or changes its orientation as it rotates, memory for the moment-to-moment spatial properties of the object must be updated. Thus, success at integrating stimuli presented at different size scales and orientations supports the idea that image–percept integration could be used to help people construct more detailed representations of dynamic environments.

REFERENCES

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

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.

Brockmole, J. R., Irwin, D. E., & Wang, R. F. (2003). The locus of spatial attention during the temporal integration of visual memories and visual percepts. *Psychonomic Bulletin and Review*, *10*, 510–515.

Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal integration of visual images and visual percepts. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 315–334.

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

Cooper, L. A., & Shepard, R. N. (1973). The time required to prepared for a rotated stimulus. *Memory and Cognition*, *1*, 246–250.

Cooper, L. A., & Shepard, R. N. (1975). Mental transformation in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 48–56.

Corballis, M., & Roldan, C. E. (1975). Detection of symmetry as a function of angular orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 221–230.

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

Di Lollo, V., & Hogben, J. H. (1987). Suppression of visible persistence as a function of spatial separation between inducing stimuli. *Perception and 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 and Psychophysics*, *55*, 373–386.

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.

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

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

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.

Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, *13*, 149–206.

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

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 mind*. 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*. Cambridge, MA: MIT Press.

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

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

Lefton, L. A. (1973). Metacontrast: A review. *Perception and 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.

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

Shepard, R. N., & Feng, C. (1972). A chronometric study of mental paper folding. *Cognitive Psychology*, *3*, 228–243.

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

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

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.

Manuscript received November 2002
Revised manuscript received April 2003