

Tests of a Temporal Theory of Attentional Binding

Steven W. Keele, Asher Cohen, Richard Ivry, Mario Liotti, and Penny Yee
University of Oregon

Different features of stimuli present in the field of view appear to be registered in different cortical maps. How, then, are the features that come from the same object bound together rather than mistakenly assembled with features coming from other simultaneously present objects? One theory supposes that an attentional mechanism intercepts input coming from particular retinal locations at a way station prior to parsing of the features from the same object. Any enhancement (or facilitation) at that stage will cause all the features from that object to be modified simultaneously in the downstream registers. The imposed temporal synchronicity serves as the essential binding cue. Five experiments provided no support for the theory. There is no tendency for synchronicity of features to cause binding unless the features come from the same location. Location, rather than temporal synchronicity, appears to be the essential cue for binding.

The various visual properties of objects in the field of view start out unified. The shape, color, size, and direction of movement all come from the same object. They come from the same place in space and they co-occur in time. When one feature disappears, typically they all disappear. But when the object is processed by the nervous system, at least some dimensions get parsed and are analyzed in differing brain regions or in different cells within the same brain region. Single-cell recordings (e.g., Cowey, 1985; Mishkin, Ungerleider, & Macko, 1983; Van Essen & Maunsell, 1983) and patterns of deficit following neurological disorder (e.g., Cowey, 1985; Posner, Walker, Friedrich, & Rafal, 1984) all suggest that differing regions in the occipital, temporal, and parietal cortex process different features emanating from the same object. This distributed processing raises a problem. Suppose that two or more objects are present in the field of view, each having a different color, different shape, different location, and the like. If one part of the brain codes color, for example, and another codes form, then how is it that later in processing it is determined which color goes with which form? What is the mechanism of reassembly?

One may question that such assembly is actually needed. It is not clear that complete separation of attributes occurs. For example, although it was once thought that Area V4 of the visual cortex was primarily a color system, recent evidence (Desimone, Schein, Moran, & Ungerleider, 1985) suggests that single cells in V4 are shape sensitive as well as color sensitive. Psychological evidence suggests, however, that assembly of attributes is in fact needed, because situations exist in which the individual attributes are perceived but are misassembled, resulting in a kind of illusion. One such demon-

stration was by Snyder (1972), who briefly presented a circular array of 12 letters and asked the subjects to name which one was red. (He also had subjects name objects that differed in other ways and obtained similar results.) Given the brief duration of exposures, subjects often made errors, but more likely than chance the errors involved naming letters that were adjacent to the correct one. The fact that errors tended to be imported from adjacent spatial positions indicates that the target features were often perceived. Otherwise, errors would be random. Even though individual features sometimes were perceived, they were nonetheless wrongly combined, indicating that they were analyzed somewhere along the processing line by differing brain systems.

More recently, Treisman and colleagues (Kahneman & Treisman, 1984; Treisman & Gelade, 1980; Treisman & Schmidt, 1982) have systematically investigated such illusory conjunctions in which the wrong features are assembled. If a target item in a display of distractor items differs from the distractors on a single feature (e.g., a red X among blue Xs), the latency to detect the target is little influenced by the number of distractor items, as though subjects survey all objects concurrently. In contrast, when the target object involves a color-form conjunction (e.g., a red X among red Os and blue Xs), reaction time to detect the target increases linearly with the number of distractors in the display, as though subjects survey each object one at a time. Treisman suggests (e.g., in Kahneman & Treisman, 1984) that sequential attention to the object from each location binds color and form. If one is prevented from attending to objects, then misassembly may occur. Such a conclusion has been supported by Prinzmetal, Presti, and Posner (1986), who showed that when attention is directed by a cue to a point in space, illusory conjunctions are less likely to occur than at other places where attention is not directed.

The psychological evidence clearly indicates, therefore, that attributes from the same object do not automatically adhere to one another but must be assembled. Just how does the assembly process operate? What are the cues that allow the nervous system to bind together the appropriate features? One possibility is that sufficient locational information resides in the cortical maps that code particular features to allow com-

The research in this article was supported by an Office of Naval Research Contract (N00014-83-K-0601) to M. I. Posner and Steven W. Keele.

Mario Liotti is now at the Institute of Human Physiology, University of Parma, Italy; Penny Yee is at the Department of Psychology, University of Washington, Seattle.

Correspondence concerning this article should be addressed to Steven W. Keele, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

mon location to serve as a binding cue. Although on the surface, common location seems reasonable as a binding cue, deeper consideration poses questions. Late-stage visual maps, which presumably fully abstract the different features, tend to have large single-cell receptive fields (e.g., Desimone et al., 1985; Ungerleider & Mishkin, 1982). The presumed reason for such loss of locational specificity is that it provides stimulus equivalence regardless of retinal locus. However, loss of locational specificity in turn raises the issue of whether the common location of features from the same object can be used to bind them together. Moreover, even if receptive fields were small, providing detailed location information, it is not clear just how the common location of features in different late-stage maps would be communicated to each other.

The psychological evidence regarding the role of location in binding is mixed. Snyder's (1972) demonstration that illusory conjunctions are more prevalent among adjacent items is consistent with locational binding. In contrast, Treisman and Schmidt (1982) did not find illusory conjunctions more prevalent among items close in space to each other. It is also known that distinctly nonlocational factors influence binding, such as gestalt principles of grouping and linguistic and semantic properties (Intraub, 1985; Prinzmetal & Millis-Wright, 1984). Given that it is not clear that common location per se is used in binding features from the same object nor precisely how a location mechanism would work, it is worth exploring alternative possibilities.

The hypothesis investigated in the present studies is that an attentional system imposes temporal contiguity between features that emanate from the same object, providing a sufficient cue for assembling features from different dimensions. To appreciate how this might work, consider the following crude analogy.

Suppose three objects are simultaneously present at different locations in the visual field, say, a red triangle, a green square, and a blue circle. Presumably these feed information into three different sensory maps. One register indicates that a red object, a green object, and a blue object exist in the external environment. Another register indicates that a triangle, a square, and a circle are present. The third register indicates that objects are present in three particular locations. The problem is how the three registers communicate to determine that it is the triangle that is red and in a particular location, and so on. Suppose now that you were to move your hand in front of your eyes. At some point it would block out one of the objects, say, the green square. At that precise moment, the green, the square, and one of the location specifications would disappear from their respective registers, and as the hand moved on, they would simultaneously reappear. Thus, although the stimuli were simultaneously present in the external world, a temporal pattern could be imposed by an agent. The temporally contiguous activation (or inhibition) of material in different registers could serve as the essential cue to bind them. Such a hypothesis has previously been proposed by von der Malsburg (1985).

Of course, the attentional mechanism is not a hand. It might be supposed, however, that at some relatively early stage in the brain, prior to the initial parsing of objects into constituent features, some attentional system exists that can

modify information coming through particular locations. It could be, for example, that in the thalamus, the various features from a given object are still carried on the same neural cells. If attention is then focused on one location in the thalamus, the features coming from that location will all be affected at the same time in subsequent cortical maps where the features are parsed even though those subsequent maps may no longer have precise location information. In essence, early locational differences become translated into temporal differences, and it is the temporal difference that serves as the binding cue.

The mechanism could be inhibitory, as in the hand analogy, so that information from the attended location could be inhibited and then released, or it could be excitatory so that the location at which attention is focused passes enhanced information. The attentional system could sweep the visual field, successively modifying inputs from the locations it passes. Such modification would have downstream affects in the separable registers for the different features, and the downstream alterations that occur at the same time could result in binding.

Crick (1984) has, in fact, proposed the thalamus as the relay station that performs the proposed attentional function. He has also suggested that the mechanism of binding across later maps is that proposed by von der Malsburg (1985). In particular, transient and simultaneous changes in activity produced by attention serve to relate the features carried in the different maps. This general view has also been discussed by Sejnowski (1986).

The present experiments were designed as one test of the hypothesis that temporal contiguity plays a critical role in the binding of features coming from the same object. The general logic is that if attention has its effects by producing temporal contiguity, then that function of attention can be mimicked by producing temporal contiguity in the external environment. In particular, features coming from different locations should tend to be erroneously conjoined when they happen at the same time. Two different paradigms were used. The first paradigm is described in Experiments 1A, 1B, and 2; the second is described in Experiment 3.

Experiments 1A and 1B

The first set of experiments made use of a paradigm developed by McLean, Broadbent, and Broadbent (1982) and further studied by Gathercole and Broadbent (1984). A series of letters was rapidly presented, one letter after another and in the same place on the presentation screen. In the McLean et al. experiments, the letters were of various colors, and in the Gathercole and Broadbent studies various colored borders surrounded the letters. In the midst of the series a single digit occurred, and the task of the subjects was to identify the color of the digit. Given the rapid presentation rate, the wrong color was sometimes reported. The interesting observation was that when the wrong color was given, it was more likely to be the color that immediately preceded (-1 condition) or followed (+1 condition) the actual target letter as opposed to two colors back (-2) or two colors ahead (+2). The error is therefore an illusory conjunction rather than a random error.

In a second version studied by McLean et al., successive letters appeared on opposing sides of a fixation point. Thus, if one item was on the left, the next was on the right, followed by another on the left, and so on. In this version of the paradigm, when a target digit occurred amidst the succession of letters, -1 and +1 items were temporally close to the target but were spatially on the opposite side of the fixation point from the target. The -2 and +2 items were further separated from the target in time but occurred in the same spatial position as the target. For this condition the majority of incorrect responses was imported from the -2 and +2 temporal positions rather than from -1 and +1 positions. The results appear to be inconsistent, therefore, with a theory of temporal binding of attributes from the same object. Instead, they suggest that features that emanate from the same location, even though at different times, may get bound. Of course, there is some temporal resolution beyond which binding will not occur even though different features come from the same location, but temporal contiguity appears not to be the key factor.

A variety of objections could be raised to the McLean et al. and Gathercole and Broadbent studies. Because alternate items appeared to the left and right of a fixation point, they projected to different brain hemispheres. Perhaps spatial contiguity is of overriding importance when it involves different hemispheres. To obviate this problem, Experiments 1A, 1B, and 2 all used a vertical arrangement in which successive items alternated above and below fixation. With the head in vertical position, such an arrangement ensures that the alternate stimuli do not go to opposing hemispheres.

Second, in neither earlier study did a masking stimulus follow each item in the series. It is possible that the residual trace of a -2 or +2 item is retained in iconic memory and merges with the target item. If this were the case, the inputs to a later decision stage might not really be temporally separated. To alleviate that problem, each stimulus in the present experiments was followed by a mask of random color patches in the hope that iconic persistence that might reduce temporal discriminability would be ameliorated.

A third issue concerns whether the phenomenon might change when the spatial separation of alternate items is reduced. Experiment 1A involved one condition in which all the items appeared at the point of fixation. In the other condition involving alternation, the centers of the stimuli were 1.6° above or below fixation. In Experiment 1B, the centers of the stimuli were only 0.56° above or below the center of the display. Comparing the two separations allows determination of whether temporal contiguity becomes more important when spatial distance becomes less distinct. It is possible that features that co-occur in time will jump space to become bound only if the distance is short.

The objective of these studies was to determine whether errors in reporting the color of the target digit embedded in a series of letters would be more likely to come from temporally contiguous or from spatially contiguous items.

Method

Subjects. Eight young adults from the subject pool of the Cognitive Laboratory at the University of Oregon participated in three sessions

of Experiment 1A. A different group of 8 subjects participated in three sessions of Experiment 1B. They were paid for their participation.

Procedure. Each individually tested subject sat in a sound-attenuated room with his or her chin on a rest. The subject's eyes were 90 cm from a color monitor. A trial consisted of a rapid series of 15 items involving 14 letters chosen from a set of 17 (A, C, E, F, G, H, J, K, M, N, P, R, U, V, W, X, and Y). The remaining item was a digit chosen from a set of six digits (2, 3, 4, 6, 7, 9). The digit appeared randomly in Position 7, 8, 9, or 10 of the series. Some letters and digits were not used to avoid confusions between the two (e.g., 8 and B are similar). The letters and digits subtended an angle of 0.7° in height and were colored black. They were superimposed on a background of white, yellow, green, blue, or red. The background was a square 1° on each side. No color was repeated within ± 2 temporal positions of the target digit, and no color was repeated in adjacent temporal positions elsewhere in the list. Together these restrictions imply that the plus 2 and minus 3 positions shared the same color and the +3 and -2 positions shared another color. Each item in the list was followed by a masking stimulus 4.7° \times 2.2° and composed of randomly arranged small patches of different colors. The different patch colors involved all the colors of the displays, including black.

There were two conditions in Experiment 1A. In the first, all the stimuli were presented at the center of the screen. In the second, stimuli alternated between two locations along the vertical meridian above and below a fixation point. The distance between the fixation point and the center of each stimulus was 1.6°. Subjects were instructed to keep fixated on the center point, but in the absence of any direct measure of fixation, subjects who detected a target at one spatial position 50% more often than at the other position were replaced in the experiment. As a result, one subject's data were discarded.

The task of all subjects was to identify the color of the background surrounding the digit. The answer was made by pressing a labeled key on the computer keyboard. Accuracy, not reaction time, was stressed. If a subject reported that the digit was not seen, the trial was replaced later. The first session of the experiment provided practice and was used to select individual exposure durations for the two test sessions. Exposure durations were selected in which it was expected that subjects would correctly name the target color between 50% and 60% of the time. In the first session, subjects were given 20 practice trials in which each stimulus was exposed for 150 ms and each was followed by a 150-ms mask. Then there were six blocks of 40 trials of which three blocks were with alternating stimuli and three involved stimuli at only one location. Each block had a different exposure setting: Stimulus and mask were each 83 ms, each 67 ms, and each 50 ms.

During each of Sessions 2 and 3, there were 240 trials divided into six blocks of 40 trials, with short rest periods between. Half the subjects received the single location condition (in which the stimulus always appeared in the center) in the first of the test sessions, and the double location condition (in which stimuli alternated above and below fixation) in the second test session. The other subjects were tested in the reverse order. Six subjects used the 50-ms stimulus/50-ms mask exposure, and 2 subjects used the 67-ms stimulus/67-ms mask exposure.

The procedure and task were identical for Experiment 1B except that in the double location condition the distance between the fixation point and the center of the stimulus positions was reduced from the 1.6° of Experiment 1A to 0.56°. With the smaller visual angle, performance improved in the double condition, necessitating different stimulus and mask durations in the second and third sessions. Three subjects used 67-ms stimulus/67-ms mask durations for the single location and 50-ms stimulus/50-ms mask for the alternating locations. The other 5 subjects used 50/50 for the single location and 33

ms for the stimulus and 50 ms for the mask for the alternating locations.

Results

Experiment A. Table 1 shows the percentage of occurrences in which the color reported came from the digit target (designated the 0 temporal position) or from the preceding or following temporal positions (designated -2, -1, +1, and +2). The correct responses (the 0 position) for the single and alternating conditions were well equated—58% and 59%, respectively. When errors were made in the single condition, more came from the -1 and +1 temporal positions, averaging 15.6%, than from the -2 and +2 positions, averaging 5.6%. Recall that in the single condition, each successive item occurred at the same point in space. For the alternating condition, in which the -2 and the +2 items were at the same place in space as the target and the -1 and +1 items were close in time but different in space, the errors averaged 8.0% for the -1 and +1 temporal positions and 12.5% for the -2 and +2 positions, a pattern reversed from that of the single condition. The interaction between single versus alternating condition and temporal position was highly significant, $F(1, 7) = 36.1, p < .001$. Averaged over the single and alternating conditions, position still achieved a main effect, $F(1, 7) = 31.4, p < .001$, favoring positions -1 and +1 over -2 and +2.

The results from Experiment 1A provide little support for the idea that temporal contiguity of features promotes binding. Instead, features that come from the same point in space, even when relatively far apart in time, tend to be bound erroneously, suggesting that spatial identity is the key binding cue. Such a conclusion also is consistent with the results of McLean et al. (1982) and Gathercole and Broadbent (1984). The changes in the present study, in which two successive stimuli in the alternating condition were projected to the same brain hemisphere and in which masks followed each stimulus to reduce any iconic image, did not alter their basic result.

Experiment 1B. It is possible, however, that temporal binding would play a more distinct role if locations were even less distinct than in Experiment 1A. To test such a possibility, the visual angle between the center of the display and the positions of the upper and lower items was reduced in Experiment 1B from 1.6° to 0.56° . The results from Experiment 1B are shown

in Table 1 along with those of Experiment 1A, and it can be seen that there was no change in the pattern of results. When the series of items all occurred in the center of the screen, more errors came from the -1 and +1 temporal positions (14.8%) than from the -2 and +2 positions (5.7%). The reverse occurred for the alternating condition, where there were 7.9% errors for the -1 and +1 positions and 13.4% errors from the -2 and +2 positions. Again, the data suggest that incorrect bindings of color to form occur primarily from items emanating from the same spatial position than from items in close temporal proximity. Reducing the discriminability of spatial location produces no tendency for temporal factors to become more important. In fact, an overall analysis of variance (ANOVA) in which Experiments 1A and 1B constituted a factor showed no modification of the interaction of temporal position with single versus alternating condition.

The only evidence that favors a role of temporal contiguity is that when data are averaged over the single and alternating conditions and over both experiments, there is a slightly higher error rate for the -1 and +1 temporal positions (11.6%) than for the -2 and +2 positions (9.3%). This difference is significant for the ANOVA in which experiment is a factor, $F(1, 14) = 26.4, p < .001$. One way of interpreting the result is that features coming from the same location tend to be integrated the closer together they occur in time. However, spatial proximity is the necessary condition for binding. Temporal proximity with spatial dissimilarity does not produce binding.

Experiment 2

In one respect the method used in Experiments 1A and 1B favors spatial binding over that of temporal binding. In the alternating condition, the -2 and +2 items occur in exactly the same place in space as the target. The -1 and +1 items, although temporally closer to the target than the -2 and +2 items, do not occur at exactly the same time as the target. Perhaps a cross-exchange of features between spatial locations would be more likely were the items at the different locations to occur simultaneously. In Experiment 2 we explored this possibility with a slight change in method. A rapid series of pairs of items occurred, one item above fixation and one below. Somewhere in the series, one of the members of a pair was a target digit. All the other items were letters. As before, a subject's task was to identify the color of the digit. Now, when an error in color identification occurred, we could ask whether it was more likely to be the color of the item in the opposing spatial location that occurred at the same time as the target or whether it was more likely to be a color that occurred at the same spatial location as the target but at a different time. With pairs of items, there were limitations in the number of successive stimuli before a color had to be repeated. Six different colors were used. The pair containing the target involved two colors, the preceding pair (-1) two other colors, and the following pair (+1) yet two others. The -2 and +2 temporal positions were not analyzed because they involved colors that were included in the critical conditions.

Table 1
Percentage of Colors Reported at Various Temporal Positions Before and After the Target Digit: Experiments 1A and 1B

Experiment and location	Position				
	-2	-1	0	+1	+2
Experiment 1A					
Single location	6	15	58	16	5
Alternating location	13	9	59	8	13
Experiment 1B					
Single location	6	16	59	13	5
Alternating location	13	8	57	8	14

Note. Because of rounding errors, some rows do not add up to 100%.

Method

Subjects. Twelve students from introductory psychology classes participated as part of a course requirement. All had normal or corrected-to-normal vision, and none were color blind. None had participated in the earlier experiments.

Task and procedure. The change to six (rather than five) background colors for the stimuli necessitated a change to a low-resolution mode because of limitations of the Apple IIe computer. To produce the correct visual angles, subjects sat 300 cm from the monitor rather than 90 cm as in the first experiments. Again a headrest controlled distance. A series of stimuli consisted of 15 pairs of items, all letters except for the target. The letters and digits were $0.60^\circ \times 0.86^\circ$ of visual angle. Each item was black and was superimposed on a rectangular background $0.76^\circ \times 1.14^\circ$. The color of the background was white, yellow, green, blue, red, or brown. The color of each item in the target-containing pair and in the preceding and following pair was different. The letters and digits were from the same sets as in Experiment 1. The center of one member of a pair was 0.76° above a fixation point, and the other was an equal distance below. The target digit was above the fixation point on half the trials and below on the other half. A mask of random color patches containing each of the different colors followed each pair and covered the entire monitor. The task was again to indicate the color of the digit by pressing an appropriate key on the computer after each series. If the subject did not see the digit, the trial was later repeated. One subject was replaced because of much greater accuracy on the upper than the lower targets, suggesting a failure of fixation to the center.

Subjects were tested in two sessions. The first was a practice session primarily for determining exposure times to produce about 50%-60% correct responses in identifying the color. The second session involved 240 trials divided into six blocks of 40 trials each. The stimulus/mask duration was 67/67 ms for 8 subjects and 83/83 ms for the remaining 4 subjects.

Results and Discussion

The mean percentage of correct identifications of the target color was 56.8%. On 6.2% of the series, subjects reported the color of the letter from the pair of items containing the target digit. This percentage was very close to the percentage of errors emanating from the preceding item in the location opposite the target (6.3%) and from the following item in the location opposite the target (4.4%). An analysis of variance showed no significant difference among the -1, 0, and +1 conditions in which the locations were all different from the target. In contrast, the percentage of errors emanating from the -1 and +1 items that shared the same spatial location as the target digit were higher—9.7% and 16.7%, respectively. Both percentages were reliably greater than error percentages emanating from the opposite location of the same pair, $t(11) = 5.14, p < .001$, for the -1 temporal positions and $t(11) = 9.20, p < .001$ for the +1 temporal positions. Also the pretarget errors from the same location as the target (the 9.7% error rate) and the posttarget, same-location errors (the 16.7% error rate) were significantly different, $t(11) = 3.5, p < .005$.

Results of Experiment 2, even more strongly than those of Experiment 1, run counter to a hypothesis of temporal binding. The likelihood of attaching a feature to a target coming from another location is no greater when the erroneous feature occurs at the same time as the target as opposed to slightly before or after the target. On the other hand, if the feature

comes from a different object occurring before or after but in the same spatial location as the target, the chances are increased that it will be erroneously attached to the target. Common location appears to be the prominent cue for binding.

It may be recalled from Experiment 1 that when a series of stimuli occurred at the same place in space, errors were imported from stimuli that were temporally nearest the target. Because this result did not occur in the present experiment for locations different from the target, it appears that temporal contiguity plays a role in binding only for items coming from the same location. Features from the same location but different times may be erroneously bound when they occur close in time, but features close in time but from different locations tend not to get bound.

Experiments 3A and 3B

Although results of the first two experiments provide no support for a theory of temporal binding of features, it is important to test the idea in a different paradigm. Conceivably, the results that favor a role of common location in binding are due to the rapid series of presentation in which stimuli replace one another in the same locations. In Experiment 3, subjects were given isolated exposures of a pair of stimuli rather than a series of stimuli. The temporal relations among the components of the stimuli were varied. In each presentation there were two letters, one above the other. Each letter was surrounded by a colored border, and the colors were different for the two letters. After the stimulus presentation a mask was presented, and then a question mark occurred adjacent to one of the two stimulus locations. The subject's task was to report the color and the letter of the indicated location. The stimulus duration was short so that sometimes errors occurred. As in the similar Treisman and Schmidt (1982) paradigm, some of the errors involved a feature error in which a color or letter that was not present in the display was reported. Other errors were conjunction errors in which either the correct color or the correct letter was reported and the other reported feature came from the uncued location in the display.

Experiment 3 contrasted the number of conjunction errors in three different conditions. In a control condition, both letters and their surrounding colors were all exposed at the same time. This established a baseline number of conjunction errors. In a second condition, the color from one location and the letter from the other location came on simultaneously for the first half of the exposure. They then went off, and immediately the other color and the other letter came on for the last half of the exposure. This condition produced temporal synchronicity between the color from one location and the form from the other location. A theory of temporal binding predicts that the number of conjunction errors should increase above baseline. In a third condition, one letter and its surrounding color were presented in the first half of the exposure, and the other letter and its surrounding color were presented in the second half of the exposure. In this case there was temporal synchronicity between the features coming from the same location. Such a manipulation should reduce the num-

ber of conjunction errors below the baseline established in the control condition. With the short exposure durations, subjects were unaware of the exact ordering of stimulus components.

In Experiment 3A, the stimuli occurred above or below a fixation point. Experiment 3B served largely as a replication of 3A, but in addition, two important changes were made. In one change the pair of stimuli, while still vertical, were closer together to reduce their spatial discriminability. As in the contrast of Experiments 1A and 1B, the issue concerned whether the cross-exchange of features having the same temporal pattern but different locations was more likely to occur when the locations were closer together. The other change was that although the two members of a pair were always one above the other, sometimes the pair was presented to the left of fixation and sometimes to the right of fixation. The purpose was to reduce any tendency of the subjects to anticipate where the target stimulus would occur and also to determine whether the role of temporal binding might differ, depending on the hemisphere to which the stimuli were projected.

Method

Subjects. Eleven young adults participated in a single session in Experiment 3A, and a different 10 subjects participated in Experiment 3B. Participation satisfied part of an introductory psychology course requirement.

Materials and procedure. At the beginning of each trial a fixation mark appeared at the center of a video monitor for 1,000 ms. At the end of that period a pair of white letters appeared, one above and one below the fixation. The four possible letters were A, Q, M, and K, and the four possible border colors were yellow, green, red, and blue. The two letters in a pair were always different as were the two colors. The height of each letter was 0.48° of visual angle, and the border was a square of 0.75° of visual angle. In Experiment 3A the vertical distance from the fixation point to the centers of the stimulus items was 1.66°. In Experiment 3B, both members of the pair appeared 1.50° to the left or to the right of fixation. The center of one item was above and the other below the horizontal position of the fixation by a distance of 0.48°. Each pair of items was followed by a masking stimulus, and then a question mark appeared adjacent to one of the former locations. The mask was full screen and involved all the colors of the stimuli, including black and white. The task was to report the letter and color that had appeared at the cued location.

There were three different exposure conditions. For the first one, which we shall call *simultaneous*, the two letters and two colors were all exposed at the same time. In contrast to the earlier experiments, all of the subjects received the same exposure durations. These were selected on the basis of pilot work to produce errors yet to allow correct responding above chance. For Experiment 3A the exposure time was 67 ms; in Experiment 3B it was 167 ms. In the second condition, which we shall call *synchronous opposite locations*, the color of one of the items and the letter of the other item were exposed for 50 ms in Experiment 3A and for 83 ms in Experiment 3B. At their termination, the other letter and color were exposed for 50 ms in Experiment 3A and for 83 ms in Experiment 3B. In the third condition, called *synchronous same location*, the letter and color from the same location were exposed for 50 ms (Experiment 3A) and for 83 ms (Experiment 3B) and were followed immediately by a 50-ms (Experiment 3A) and 83-ms (Experiment 3B) exposure of the letter and color from the other location. The longer exposures for Experiment 3B were needed to compensate for the additional uncertainty of spatial position. At the end of the stimulus display, a mask of

colored patches covered the entire video screen for 100 ms, and then the question mark appeared, indicating which item to report.

After a practice block of 20 trials with the simultaneous condition only, subjects received eight more blocks of 48 trials, with short rest intervals between each. The three types of displays were randomly intermixed within each set of 48 trials.

Results and Discussion

Experiment 3A. The mean percentage of times in which both features of the stimulus were correctly reported were relatively similar for the three exposure conditions—57% for the simultaneous condition, 54% for the condition of synchronous opposite locations, and 53% for the condition of synchronous same locations. Our main interest, however, is in the breakdown into feature errors and errors of conjunction. Feature errors are cases in which one or both of the reported features did not occur in the display. Conjunction errors refer to a situation where one feature is from the cued location and the other from the uncued location. For the simultaneous condition there were 32% feature errors and 11% conjunction errors. For the condition of synchronous opposite locations, there were 34% feature errors and 11% conjunction errors. And for the case of synchronous same locations, there were 34% feature errors and 11% conjunction errors. These two types of errors were virtually identical in proportion across all three conditions; the small differences were not statistically reliable. In addition to conjunction and feature errors, subjects sometimes reported both features from the uncued location. These errors, which were not counted as either feature or conjunction errors, averaged 2% and did not differ among exposure conditions.

These results provide no evidence that features from different locations that occur at the same time tend to conjoin. In fact, they provide no evidence for true conjunction errors beyond what would be expected by chance, given the frequency of feature errors. Consider the case in which one of the cued features was correctly reported and the other was in error. Because there were twice as many possible features that were not displayed at all on the trial as in the displayed but uncued location, there should have been twice as many feature errors as conjunction errors if subjects simply guessed at the feature they did not perceive. Averaging the data over the three display conditions yielded 11% conjunction errors. The cases with one feature error (eliminating the case in which there were two feature errors) constituted 25%, which is slightly more than twice as great as the conjunction error rate. Thus, the percentage of conjunction errors in this experiment was not greater than would be predicted by guessing. It is not clear why no conjunction errors occurred in Experiment 3A, but the different arrangements in Experiment 3B did yield clear conjunction errors, allowing a stronger test of the temporal synchronicity hypothesis. It may be that conjunction errors do not rise above chance unless exposure conditions are sufficient to allow a larger percent correct.

Experiment 3B. The purpose of this experiment was to explore whether conjoining of features from differing locations would be more likely under conditions of temporal synchronicity if the locations were less distinct. The two

locations at which stimuli appeared were about as close together as possible without any overlap of the target items. In addition, the experiment allowed a test of whether there would be differences in outcome for stimuli presented to the left or right visual field and hence to the right or left brain hemisphere.

Table 2 shows the results for the three exposure conditions and the field of exposure. Too long an exposure time was selected for the simultaneous condition so that it had a larger percentage of correct reports of both features. However, the two other conditions, which are the critical ones, produced nearly identical percentages of correct responses. The question of central interest, then, is whether the case in which synchronicity occurred among features from differing locations produced more conjunction errors (and fewer feature errors because the total of correct responses is approximately equated) than did the case in which synchronicity was among the components from the same location. In fact, the percentage of errors of the two types was close to the same in the two exposure conditions but tended to be in the wrong order from the prediction of the temporal binding theory. That is, there were slightly *fewer* conjunction errors in the case in which a feature from one location co-occurred with a feature of the other dimension from the other location. Because the tendency is opposite that predicted, there is no need for a statistical comparison. In addition, error rates in which both reported features were from the uncued location do not differ between the different exposure conditions.

The mean rate of conjunction errors was 12% averaged over the synchronous opposite and the synchronous same locations. The mean rate of feature errors where only one of the features was in error and the other was correct averaged 17%. Because the percentage of conjunction errors was more than 50% of the percentage of feature errors, it appears that more illusory conjunctions occur than can be attributed simply to guessing of the features not perceived. Despite the presence of illusory conjunctions, their frequency is not altered by the different temporal patterns of exposure.

Although subjects' responses were more accurate for stimuli presented to the right visual field than for those to the left,

the pattern of results among the three exposure conditions was not affected by visual field. There was no tendency for temporal synchronicity to facilitate conjunctions in one field as opposed to the other.

Experiment 3B was thus consistent with Experiment 3A in failing to provide any evidence that temporal concurrence of features from different dimensions in different parts of the visual field causes them to become conjoined.

General Discussion

The three experiments in this study were all designed to test a hypothesis of the manner in which attention binds together features (e.g., color and form) that come from the same object. Such a binding process appears necessary on both physiological and psychological grounds. On physiological grounds, different portions of the cortex appear to predominantly code differing features of the stimulus. Thus, the conscious recognition that certain features belong to the same object seems to require a reintegration. Psychologically, a binding process appears necessary because under impoverished conditions, people may perceive the features but misassemble them, producing illusions. Such errors are reduced for objects in locations toward which attention is directed. An important issue in the analysis of attention concerns the mechanism by which binding occurs.

Crick (1984), von der Malsburg (1985), Sejnowski (1986), and we ourselves have suggested that an attentional mechanism intercepts visual information at a way station prior to its parsing into separate features. Attention modifies the input coming from the currently attended location either by enhancing or inhibiting the input. Such modification causes temporal synchronicity at the downstream registers of all the features coming from the attended object. In essence, location information becomes converted to a more useful temporal information. According to the hypothesis, it is the temporal contiguity that causes the binding, not the common location of origin.

None of the experiments in the present study provide any support for such a theory of temporal binding. If temporal binding occurs, then it would be expected that features such as color and form coming at the same time from different locations would tend to be combined. There is, however, no tendency for more false conjunctions from simultaneous features than for ones that do not co-occur.

A weakened version of the temporal theory is that multiple cues contribute to binding, with time being one cue and location another. In such a view one might expect the importance of space and time to vary, depending on their salience. However, the present studies provide no evidence for that weakened theory either. There was no tendency for temporally based binding to increase either as locations became less distinct or as different features became more synchronized.

If temporal contiguity is not responsible for binding, then what is responsible? An alternative is that despite reservations based on large receptive fields and the lack of an explicit model, location per se is the binding agent, a possibility supported in work by Nissen (1985). When objects of different shape and color are briefly presented at different locations in

Table 2
Mean Proportion of Responses Correct and in Error:
Experiment 3B

Condition	Correct	Conjunction error	Feature error	Uncued pair
Right visual field				
Simultaneous	.88	.03	.09	.00
Synch. opp. loc.	.71	.12	.18	.02
Synch. same loc.	.72	.15	.13	.02
Left visual field				
Simultaneous	.81	.05	.15	.00
Synch. opp. loc.	.66	.12	.21	.01
Synch. same loc.	.67	.13	.19	.01

Note. Synch. opp. loc. = synchronous opposite location; synch. same loc. = synchronous same location. Because of rounding errors, some rows do not add up to 1.00.

the visual field, the abilities to detect shape and color are independent of one another, suggesting that they are registered in different systems. When subjects are asked to name the location and shape of a precued color, the conditional probability of perceiving the shape is much lower if the location is not perceived. In other words, knowing that a shape goes with a particular color appears to depend heavily on knowing the location of the color. This suggests, in turn, that locational coordinates are critical.

The current experiments, particularly Experiment 2, are consistent with Nissen's conclusion of location binding. In particular, features that belong to different objects are more likely to get erroneously bound together if they come from the same location even if they are presented at different points in time.

Although the current experiments provide no evidence for temporal binding, there is at least one possible way to salvage a variant of the theory. The logic of the studies has been that the hypothesized function of attention, producing a temporal pattern, can be mimicked by externally imposing a temporal pattern. It is possible that temporal binding is a viable mechanism, but a safety exists to prevent accidental binding of simultaneous features from different locations. Binding may occur *only* if the attentional mechanism itself is responsible for producing the temporal pattern. In other words, the temporal pattern may have to be internally generated. The present data pose a problem for this notion as well but do not clearly rule it out. Consider Experiment 3. Stimuli are presented at two different locations, but in one condition the components of a stimulus at a particular location occur at different times. Conceivably, some kind of iconic image is built that effectively erases the temporal differences of the components, and then the attentional mechanism operates to reimpose its own temporal pattern in such a way that synchronicity occurs between features coming from the same place in space. The joint operation of attention concurrent with the temporal pattern it produces may result in binding.

The problem with the notion comes when it is applied to Experiments 1 and 2. Suppose stimuli occurring at slightly different points in time are compressed into an icon so that the original temporal tags are lost. Then attention operates to produce a new temporal pattern. This would result in a loss of temporal discrimination of items adjacent in time and should, in turn, produce a very large number of conjoining errors between such temporally adjacent items. However, the data show a tendency, albeit imperfect, for stimuli occurring at separate points in time to be recognized as indeed being separate stimuli. The original temporal information appears retained to a large degree, and it helps to prevent confusions of two items coming at different times through the same location.

Overall, therefore, the current data appear most consistent with the following conclusion: Both color information and form information retain locational information in their respective registers. In addition, the information is tagged in time of occurrence. Attention binds color and form when they come from the same location and in approximately the same time frame. Time per se is not the binding agent. Rather, location is primary, but it must also take into account the

time of occurrence. Otherwise, there would be no ability to differentiate objects that occur at different times but at the same place.

Two issues remain that must be addressed, given the conclusion of location-based binding. One is the notable failure of Treisman and Schmidt (1982) to find distance between objects to influence the frequency with which their features exchange. Much as Snyder (1972) found, one might expect that if common location was instrumental in binding, close-by objects would be more likely to produce illusory conjunctions. This issue has been addressed by Cohen (1987). He has shown that the focus of attention can be broad or narrow. Within the focus of attention the frequency of illusory conjunctions does not depend on the distance between items. Illusory conjunctions outside the focus depend on distance. This conclusion explains the Treisman and Schmidt results, because in their study all the items that could contribute to illusory conjunctions were within the focus of attention.

The second remaining issue concerns just how location-based binding works. We provide no complete mechanism here. However, Cohen's work (1987) suggests that location is not a fixed tag that accompanies other features. Instead, localization can be narrow or broad, depending on the focus of attention. Conjoining of features in some way occurs only for those features that appear within the focus and within the same time frame.

References

- Cohen, A. (1987). *Illusory conjunctions as a function of focus of attention and distance between objects*. Doctoral dissertation, University of Oregon, Eugene.
- Crick, F. H. C. (1984). The function of the thalamic reticular spotlight: The searchlight hypothesis. *Proceedings of the National Academy of Sciences, USA*, 81, 4586-4590.
- Cowey, A. (1985). Aspects of cortical organization related to selective attention and selective impairments of visual perception: A tutorial review. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 41-62). Hillsdale, NJ: Erlbaum.
- Desimone, R., Schein, S. J., Moran, J., & Ungerleider, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Research*, 25, 441-452.
- Gathercole, S. E., & Broadbent, D. E. (1984). Combining attributes in specified and categorized target search: Further evidence for strategic differences. *Memory & Cognition*, 12, 329-337.
- Intraub, H. (1985). Visual dissociation: An illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 431-442.
- Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), *Varieties of attention* (pp. 29-62). New York: Academic Press.
- McLean, J. P., Broadbent, D. E., & Broadbent, M. H. P. (1982). Combining attributes in rapid serial visual presentation tasks. *Quarterly Journal of Experimental Psychology*, 35A, 171-186.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience*, 6, 414-417.
- Nissen, M. J. (1985). Accessing features and objects: Is location special? In M. I. Posner & O. S. Marin (Eds.), *Attention and Performance XI* (pp. 205-220). Hillsdale, NJ: Erlbaum.
- Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal lobe injury on covert orienting of visual attention.