

**Title:** Temporal contextual cueing of visual attention

**Journal:** *in press at Journal of Experimental Psychology: Learning, Memory, and Cognition*

**Authors:**

Ingrid R. Olson  
Yale University

Marvin M. Chun  
Vanderbilt University

**Running Head:** Temporal Contextual Cueing

**Number of characters and spaces in abstract:** 943

\*address correspondence to:

Ingrid R. Olson  
Department of Diagnostic Radiology,  
Yale School of Medicine  
333 Cedar Street  
Fitkin Bldg, Rm F14  
New Haven, CT 06510

**telephone:** (203) 785-5052

**fax:** (203) 785-6534

**e-mail:** iolson@boreas.med.yale.edu

#### Abstract

Previous research has shown how spatial attention is guided to a target location, but little is understood how attention is allocated to an event in time. Here, we introduce a paradigm to manipulate the sequential structure of visual events independent of responses. We asked whether this “temporal context” could be implicitly learned and used to guide attention to a relative point in time and/or location in space. Experiments show that sequentially structured event durations, event identities, and spatio-temporal event sequences can guide attention to a point in time, as well as to a target event's identity and location. Cueing was found to rely heavily on the element immediately preceding the target, although cueing from earlier items was also evident. Learning was implicit in all cases. These results show that the sequential structure of the visual world plays an important role in guiding visual attention to target events.

---

Sequential processing of stimuli plays a critical part in most of human cognition. Auditory processing is defined by its reliance on sequential processing, but the involvement of sequential processing in vision is less obvious. Besides basic motion perception, navigation and reading are examples of visual activities that involve viewing an ordered series of stimuli (Wallis & Bulthoff, 1999; Lashley, 1951). Each stimulus is related to the

previous and future stimulus. The sequential order of events creates a temporal context or structure. Does the brain ignore this prevalent and invariant information or is it extracted and learned to optimize visual behavior? It is clear that sequential structure contains salient information about future stimuli. Such temporal context may help create expectancies about what may come next, and when it will happen. In other words, temporal context may direct attention to

important events as they unfold over time. The importance of sequential order has been highly studied in relationship to memory (e.g. Brown, Hulme, & Preece, 2000; Estes, 1985; Lewandowsky & Murdock, 1989), language (e.g. Chomsky, 1957; Elman, 1991; Reber, 1988; Saffran, Aslin, & Newport, 1996), tones (Saffran, Johnson, Aslin, & Newport, 1999), and motor tasks (Doeringer & Hogan, 1998; Jordan, 1995; 1997; Nissen & Bullemer, 1987; Stadler, 1995) but less is known about how serial or sequential order of events affects visual attention.

Various lines of research hint at how temporal attention may operate. One starting point is to examine whether temporal attention is similar to spatial attention. Decades of research have shown that responses are facilitated when attention is directed to a location in space (see Chun & Wolfe, 2000, for a review), either by stimulus-driven cues (Eriksen & Hoffman, 1972; Posner, Synder, & Davidson, 1980; Yantis & Jonides, 1996) or by instruction (Posner & Cohen, 1984). Are there temporal analogues to these spatial cues? Coull and Nobre (1998) created an analog to the Posner cueing task (Posner et al., 1980) in which they manipulated participants' expectations of where or when a target stimulus would appear in the display. The central cue was a compound stimulus, consisting of diamond within a circle. The cue directed participants to attend either to the position of the target (left or right), by highlighting one side of the diamond, or to the timing of the target's appearance (300 or 1500 ms after cue onset) by modifying the circle. Findings showed that participants' performance was facilitated by both valid spatial and valid temporal cues. Thus, temporal

## Temporal Contextual Cueing

attention can benefit from explicit cues in a similar manner as spatial attention (also see Kingstone, 1992). This type of cueing undoubtedly plays an important role in guiding attention to a point in time. However, an important ecological factor is missing from these past studies.

Temporal information in vision is comprised of a series of events, forming a temporal context. We rarely experience stimuli devoid of spatial or temporal context. Scenes are viewed and remembered relative to other scenes. Within each scene and between scenes, meaningful regularities exist among objects, locations, and times, forming spatial and temporal structures. The invariant structure found in the time domain of visual information may tune vision to points in time, as well as provide information for other behaviors. Lashley (1951) discussed how temporal sequences were important in nearly every behavior imaginable. For instance, competent speaking and walking rely upon well-formed temporal structure.

More recently, it has been suggested that temporal structure plays an important role in learning to recognize objects at unusual, or non-canonical viewpoints. As we approach, walk around, or haptically manipulate an object, we see it undergo a sequence of viewpoint transformations. Wallis and Rolls (1997) suggested that the temporal regularities found in an invariantly ordered sequence of views modifies object representations to make them viewpoint invariant (but see Harman & Humphrey, 1999). Computer simulations have been used successfully to model viewpoint invariance (Wallis & Baddeley, 1997). This suggestion also has some psychophysical support. For instance, Wallis (1998) exposed participants to three sequences of

different faces at different viewpoints such as  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $+45^\circ$ , and  $+90^\circ$ . The viewing impression was of seeing a head rotate smoothly from left to right profile. The temporal order hypothesis predicts that exposure to invariant sequential presentations will cause faces to be inadvertently associated. If this is true, then participants should have learned associations between particular views of different faces. Following the experiment, participants performed a same/different judgment on briefly presented faces. More discrimination errors were made for faces that had been shown in the same temporal sequence during training, than for faces which appeared in different temporal sequences during training. The results suggest that temporal order may be an important variable in recognition.

These results are interesting but leave some questions unanswered (Wallis & Bulthoff, 1999). First, it is not clear what aspect of the stimulus was learned. Second, it is not clear how the temporal sequences were learned. Most people do not consciously attend to such information, or make attempts to encode it. Were participants aware of the temporal structure and consciously trying to encode it? Third, it would be interesting to understand how the temporal and spatial dimensions interacted.

It is possible that temporal invariant structure, or "temporal context", guides vision in a manner analogous to how spatial context influences perception. The role of invariant spatial context has recently been investigated using a new paradigm, termed *contextual cueing* (Chun, 2000; Chun & Jiang, 1998). In this task, participants perform a visual search task over a display of rotated T's (target) and

### Temporal Contextual Cueing

L's (distractors). The participant locates the target as quickly as possible and makes a response. Unknown to them, certain display configurations repeat over time. Chun and Jiang (1998) showed that when a configuration was associated with a target in a specific spatial location, this context information improved search performance. This learning occurred for repeated configurations that were separated by several novel trials. These results suggest that invariant spatial context information can be learned to guide attention to important locations in scenes, a process called contextual cueing. Unlike previous findings, contextual cueing of attention occurred in the absence of explicit learning instructions or bottom-up stimulus factors. These results are important since they suggest a natural mechanism for orienting attention to a location in space involves sensitivity to invariant properties of the spatial world.

In this paper we examined whether temporal context can guide attention to points in time, just as spatial context guides attention to locations in space. In other words, we investigated a temporal analogue to spatial contextual cueing. If temporal attention operates in a similar manner as spatial attention, a learned sequence of visual events may guide visual processing to important points in time, such as the impending appearance of a target. We designed a new paradigm in which the temporal or sequential structure is defined over visual events. This paradigm, termed temporal contextual cueing, involves searching for a target in a rapidly changing stream of distractors. The sequential organization of the distractors preceding the target contains invariant temporal information.

The task for all of the experiments is visual search for targets appearing in sequences of visual stimuli. The experiments all follow the same basic design. Each session starts with a Training Phase, followed by a Testing Phase and then a Recognition Phase. In the Training Phase, participants are presented with a set of visual stimulus sequences, which are repeated from block to block. Over time, there should be a general facilitation of response time (RT). In the Testing Phase, participants are presented with a set of novel stimulus sequences. If participants learn the invariant sequences during training, and if this sequence information guided search, then performance should be worse in the Testing Phase, compared to the later portions of the Training Phase. Finally, in the Recognition Phase, participants are asked to determine whether they could explicitly recognize stimuli sequences as having been presented throughout the training session.

Sequence learning has been studied in the past using the serial reaction time (SRT) paradigm (Nissen & Bullemer, 1987). In the SRT task, each stimulus is presented individually at one of four screen locations. Upon seeing each stimulus, participants make a keypress corresponding to the target's location, and then the next trial begins. In the experimental condition, the spatio-temporal sequence of lights on the computer screen follows an invariant pattern across trials, whereas the pattern is random in the control condition. Over time, participants typically respond faster to the invariant pattern. In addition, participants are generally unaware of the presence of an invariant sequence, leading some researchers to conclude that SRT-type learning is

Temporal Contextual Cueing implicit (e.g., Lewicki, Czyzewska, & Hoffman, 1987; Nissen & Bullemer, 1987). However, some participants become aware of various aspects of the repeating pattern (Cleeremans & McClelland, 1991; Nissen & Bullemer, 1987; Perruchet & Amorim, 1992; Willingham, 1999). In any case, awareness does not seem necessary for learning to occur.

The SRT task has the advantage of being quite simple, but for our purposes, it has the disadvantage of a strong motor component (Hoffmann & Sebold, 1996; Willingham, 1999). For instance, each stimulus in a sequence is associated with a response. In addition, invariant sequences of stimuli are typically correlated with invariant sequences of motor responses. In contrast, we are interested in associative sequences of visual events, not motor responses. A response is not required for every item within our invariant perceptual sequences. In addition, motor responses are decorrelated from the invariant sequences. Thus, our temporal contextual cueing paradigm is tailored for studying effects of temporal structure in perception.

We investigated three different types of visual temporal structures. Accordingly, the paper is divided into three sections, each one considering a different type of temporal structure. In Section 1, we created temporal sequences of event durations. Each sequence contained stimuli of varying durations, forming a visual rhythm. For example, the first stimulus in a sequence would appear for 80 ms, the second stimulus would appear for 1066 ms, the third stimulus would appear for 266 ms, and so on. Several sequences were generated and repeated across blocks, making them invariant. Each sequence

was associated with a target appearing at the end of the sequence. If subjects were sensitive to the invariant visual rhythm (i.e., the temporal structure or context), then this would cue them to the onset of the target, facilitating detection.

In Section 2, the temporal sequences were defined by invariant event identities. Unlike the temporal manipulation in Section 1, in Section 2 temporal structure is defined by the serial order of stimuli. For example, "B T M G X ..." form a temporal sequence that can be distinguished from "G M T X B ...". Thus the invariant sequence is found in the order of stimulus events, which is another form of temporal structure (e.g. ordinal timing). Finally in Section 3, the temporal sequences contained invariant spatio-temporal information. Each trial contained a sequence of stimuli that appeared in locations that changed from frame to frame. Thus the sequences were again defined by serial order of stimuli.

## **Section 1. Temporal contextual cueing from invariant rhythm of event durations**

### **Experiment 1A**

Temporal structure can be divided into two time-based variables: a sequential or ordinal variable, and a duration or rhythm based variable. The former variable refers to the order in which stimuli occur. The later variable, duration, is a fundamental stimulus variable, theorized to be part of a counting or frequency estimation mechanism in animals, assisting in foraging and predation behaviors (Gallistel, 1990). There is also evidence that timing and rhythm affect attending (e.g., Jones & Yee, 1997; Klein & Jones, 1996; Large & Jones, 1999). Jones and

### **Temporal Contextual Cueing**

colleagues have proposed a model of dynamic attention in which rhythm and rate can focus auditory and visual attention to important time points. Because duration is a basic time variable, we considered this a good starting point for investigations of the role of temporal order in attending. Thus the experiments in this section manipulate both duration and sequential order.

We hypothesized that the sequential order of event durations could direct attention to an important point in time. This would be shown by speeded responses to targets following invariantly ordered event durations, a fixed visual rhythm. This was assessed in the Training Phase of each experiment. To ensure that the speeded responses were due to sequence-specific learning, and not a general practice effect, we tested participants on randomly ordered event durations. We expected that this would slow responses to targets to baseline level. This was assessed in the Testing Phase of each experiment. Last, we hypothesized that participants would encode the temporal sequences without effort or awareness. This was assessed by a recognition test given at the end of each experiment. If learning was implicit, hit and false-positive rates would be equivalent for learned and novel stimuli. Figure 1 illustrates the design of a sample trial.

More specifically, during the Training Phase, participants searched for a target in a stream of letters. The letters preceding the target appeared for different durations, forming a duration-sequence, which was repeated many times over the course of the experiment. These sequences are termed "invariant sequences" and constitute the temporal context. For instance, an invariant

sequence on trial 1 could be: R (80 ms), T (666 ms), Y (1066 ms), X (80 ms) . . . Later, on trial 20 the same invariant sequence might come up as: B (80 ms), Y (666 ms), C (1066 ms), K (80 ms). Each temporal sequence predicted the serial position of a randomly chosen target. The target immediately followed each invariant sequence. Participants looked for the target, and entered their response into the keyboard, as quickly as possible.

During the Testing Phase, participants did the same task. Unknown to them, the sequences had been manipulated so that the invariant structure of stimulus durations was disrupted, that is, each sequence contained a random ordering of both durations and letters, different from those used in training. We compared performance at the end of training to performance in the testing condition, to assess sequence-specific learning.

### Method

Participants. In all the experiments reported in this paper, participants were either paid volunteers or introductory psychology students participating in this experiment for payment or credit. All participants reported normal or corrected-to-normal vision. None were aware of the purpose of the experiment. Ten participants took part in Experiment 1A.

### Design.

Conditions. On every trial in the Practice, Training and Testing Phases, participants searched for a predefined target, either an X or K. A target was present on every trial. The experiment began with a 16-trial practice session. This was followed by 12, 16-trial blocks of Training Phase (e.g., eight invariant sequences crossed with the two

Temporal Contextual Cueing targets to create 16 trials per block). This was followed by four, 16-trial Testing Phase blocks. The experiment ended with a recognition test.

Stimulus Materials- Training and Testing Phases. All distractor and target stimuli consisted of black alphabetic letters presented at central fixation. The letters were randomly chosen at the beginning of each trial. The serial order of the letter identities was variable and was randomized for each trial. The invariant structure was found in the invariant “visual rhythm” established by event durations for each frame in the temporal sequence. Five durations were used to form the invariant sequences: 80, 266, 666, 1066, and 1600 ms. The target was always presented for 80 ms.

Each trial consisted of 15 items, one target and 14 distractors, presented in a sequential stream. Invariant sequences are defined as those that were repeated throughout the Training Phases of the experiment. The invariant sequence formed only a portion of the 15 items presented each trial. The invariant sequence always commenced with the first letter of each trial and continued until the target appeared. Eight invariant sequences, from three to ten letters in length, were generated for each participant. The different sequence lengths insured that the position of the target was unpredictable across trials, appearing anywhere from the fourth to the eleventh frames. Each invariant sequence was associated with a target in a particular serial position, for example, the shortest invariant sequence (three) was associated with a target in the fourth position.

The last portion of the sequence, following the target, was a letter sequence in which both letter identities

and letter durations were randomized. This served to mask the target and to equate total sequence length at 15 letters. Each letter in the tail-end sequence appeared for 100 ms. This served to shorten the duration of each trial. Although this faster presentation rate may have served as a post-target cue, signaling that the target had occurred, it was equivalent for all sequences used in training and testing, and hence this cue should not have unevenly affected the results.

During the Testing Phase, participants were exposed to sequences that were identical to the sequences in the Training Phase except that the order of the event durations was randomized on each trial. Thus the sequences contained no invariant information. The Testing Phase onset was invisible to participants. As before, all sequences were 3 to 10 items long, targets were in the fourth through eleventh positions, and were followed by a tail-end sequence to equate sequence lengths. Frequency of serial position was equated to the Training Phase. The same event durations used for the Training Phase sequences were used for Testing Phase sequences. However, the overall duration of each sequence in the Testing phase may have differed from those in the Training Phase because the durations were randomly sampled anew for the test sequences. For instance, a four item invariant sequence during the Training Phase may have been: 80, 266, 1066, 666. During the Testing Phase, a four item sequence may have had a longer total duration, e.g. 1066, 80, 266, 1600. This is discussed in greater detail in the Discussion.

Stimulus Materials- Recognition Phase. Stimuli consisted of interleaved trials of invariant sequences,

Temporal Contextual Cueing used in the Training Phase, and novel sequences. Eight invariant sequences were presented and eight new sequences were presented. Instead of searching for a target, participants were instructed to respond “yes” if they thought they had seen this particular configuration in the earlier visual search blocks or “no” if they didn’t recognize this configuration. Accuracy was recorded.

#### Procedure

The experiment was carried out in normal room illumination held constant for all participants. The experiment was self-paced. The participants’ task was to monitor the visual stream and report which target letter appeared in the stream, entering their response, either “X” or “K” as quickly as possible into the computer using the keyboard. The participant began each trial by pressing the space bar on the computer keyboard. A fixation dot appeared for 500 ms and then went off for 13 ms. The stream of stimuli then appeared successively without interstimulus blanks at the same location for durations of 80-1600 ms each. The sequence was followed by a fixation dot. Participants were encouraged to enter their response at the earliest time possible, even if the trial was not yet complete. Reaction times were benchmarked to the target onset in this and all other described experiments. Positive feedback was given for all trials in practice, training, and Testing Phases, in the form of high-pitched beeps, negative feedback in the form of a low-pitched buzz in this and all other experiments.

Participants were informed about the purposes and manipulations in the experiment prior to the onset of the Recognition Phase. Specifically they were told "You may not have noticed,

but there was a temporal pattern to some of the letter sequences you saw. This temporal pattern was in the form of a visual rhythm. In this next section we want to see if you can recognize any of these rhythms." Participants were not told that "junk" sequences followed each invariant sequence because this may have modified recognition strategies and harmed possible explicit performance. Participants were encouraged to ask questions about the invariant sequences and /or the manipulation. Additionally, participants were informed that half of the recognition trials were composed of completely novel sequences. Responses were unspeeded and no feedback was given. The Recognition Phase was followed by debriefing.

Apparatus. All of the experiments were conducted on a Macintosh computer. The software used for running the experiments was MacProbe version 1.6.8 (Hunt, 1994). The letter stimuli were presented in Helvetica font, point size 24, on a 17" monitor. The stimuli were viewed from an average distance of 40 cm. The background was a uniform gray field.

### Results

In this and all subsequent experiments, performance measures are based only on trials in which the target was correctly identified, and RTs above 4000 ms are discarded (less than 1% of trials). RT data were collapsed across all eight invariant sequences. The statistics presented in this section refer to the block manipulation (e.g. the change in performance across time). Mean correct report performance was 87.5%. Block data were grouped into larger units called epochs to increase power. Each epoch contained 4 blocks. The means of correct RTs were computed for each epoch and condition. Mean accuracy by

Temporal Contextual Cueing epoch for all experiments is presented in Table 1. There were no significant differences among the accuracy measures at different epochs,  $F(3,27) = 2.03, n.s.$  in Experiment 1. This indicates that accuracy was relatively high throughout the experiment and was unaffected by the duration sequence manipulation.

Figure 2 presents the mean RT for each epoch. A repeated measures analysis of variance (ANOVA) of RT showed that RTs became faster during the Training Phase,  $F(2, 18) = 8.96, p < .002$ . A second learning effect was assessed by conducting a t-test between the last epoch of the Training Phase (epoch 3) and the testing condition epoch (epoch 4). This revealed a statistically significant effect of condition  $t(9) = 2.146, p < .041$ . This suggests that much of the RT decline over the course of training was due to a benefit derived from encoding the specific invariant sequences.

Explicit Recognition Test. Recognition performance was assessed at the end of the experiment, and in all other experiments reported in this paper. In all experiments we compared hit rate and false positive rate in an ANOVA. In this and all other experiments, hit rate was calculated as the number of times the participant responded "yes" to an invariant sequence divided by the total number of invariant-sequence trials in the Recognition Phase (e.g. 8). False positive rate was calculated as the number of times the participant responded "yes" incorrectly, to a novel sequence, divided by the total number of novel trials (e.g. 8). Mean accuracy was calculated as the number of hits plus the

number of correct-rejections (e.g. the number of times the participant responded “no” correctly, to a novel sequence) divided by the total number of trials. In this experiment, mean accuracy was 50%. There was no difference between the mean hit rate of 52% and the mean false-positive rate of 52%.

### Discussion

The results from Experiment 1A show that invariant temporal context facilitates participants' performance relative to a control condition in which temporal sequences were randomized. Thus, temporal context or structure in the form of invariant event durations was learned and cued when a target was to occur. The absence of an invariant pattern in the randomized sequences of the Testing Phase impaired performance. This suggests that participants' performance was enhanced by sequence-specific information and not solely due to general motor or skill learning. In sum, invariant event duration sequences can be learned and such temporal context guides attention to a point in time.

At the end of the experiment, participants performed a yes/no recognition task. Explicit recognition performance was at chance, leading us to conclude that learning was implicit. This is important for two reasons. First, much of the current interest and debate in sequence learning using the SRT task has centered around the implicit learning/explicit learning distinction (e.g., Clegg, DiGirolamo, and Keele, 1998; Nissen & Bullemer, 1987; Perruchet and Amorim, 1992; Shanks & St. John, 1994; Willingham, Greenley, and Bardona, 1993). Although some researchers report that normal and amnesic participants are unaware of the invariant pattern, (e.g., Lewicki, Hill, &

### Temporal Contextual Cueing

Bizot, 1988; Nissen & Bullemer, 1987), some participants can become aware of the repeating pattern (Cleeremans & McClelland, 1991; Nissen & Bullemer, 1987; Perruchet & Amorim, 1992) leading to the current debate about whether or not this type of learning is really implicit (Shanks & St. John, 1994).

The SRT task has a strong motor component that requires a response to every stimulus (Hoffmann & Sebald, 1996; Willingham, 1999). It is possible that motor sequence learning or stimulus-response learning, more often leads to explicit awareness. This may explain why investigators studying the SRT often find explicit awareness, while we did not in a similarly sequential, though primarily visual task.

Second, because we are interested in linking these findings to ecological visual problems, it is important that learning arises by simply interacting with the environment. Seeing is so immediate and easy that it seems unlikely that we consciously try to seek statistical patterns in the input. By this logic most forms of visual learning that guide our perception and interactions with the environment should occur in an implicit or incidental manner (Chun, 2000). This experiment demonstrates that temporal structure can be implicitly learned (see also Saffran et al., 1996, 1999) and used adaptively to allocate attention to a place in time. However, it should be noted that it is possible that explicit knowledge and performance were attenuated in Experiment 1A by the fact that it was probed after inclusion of many blocks of random sequences.

This experiment leaves open the question of what is being learned. There are two possibilities. First, participants could be learning a visual rhythm in

which the order of durations is the important predictor of target appearance. Alternatively, participants may have learned a set of eight absolute sequence durations. For instance, one plausible strategy for performing well on this task would be to time target onset in absolute time from the first stimulus onset. An astute observer would note that there were eight most probable times after sequence onset at which targets occurred. This type of learning would bypass the need to pay attention to the order of the individual durations since a stop-watch type mechanism would only care about the absolute time from trial onset. Since absolute time was not controlled in the test phase, we cannot rule out this proposal. We examined this alternative hypothesis in Experiment 1B.

### **Experiment 1B**

In Experiment 1B we tested the hypothesis that sequences of durations, and not absolute times, were directing attention to a relative point in time, in the paradigm used in Experiment 1A. This was examined by controlling the absolute time of each invariant sequence. If temporal context still directed attention to the target event when absolute overall time is controlled, it can be concluded that sequences of durations were being learned.

We also provide further tests of the specificity of learning and cueing in two types of interleaved test trials. In Test 1, the order of the durations used in the Training Phase were randomized. By comparing RTs in the Training Phase to those in Test 1, we could assess sequence-specific learning, while controlling for the absolute duration of the sequences. In Test 2, we tested the specificity of cueing. The learned sequences may cue the onset of the

Temporal Contextual Cueing target in a precise manner or they may simply be generating a vague expectancy that a target is likely to occur, say, within the next few hundred milliseconds. In other words, if participants learn that the invariant sequence immediately precedes the target, then cueing should be disrupted when extra elements are inserted between the invariant sequence and the target. If learning produced a vague expectancy of target onset, then there should be little disruption by intervening elements. Thus, Test 2 will help illuminate the specificity of the relationship between predictive sequences and the temporal onset of associated targets.

The method used in Experiment 1B is similar to that used in SRT tasks (e.g. Nissen & Bullemer, 1987). Only one sequence is learned and the beginning and end of the sequence are obscured by random "junk" sequences. Each participant was exposed to a unique invariant sequence. Unlike Experiment 1A, the invariant sequence never begins with the first frame. Otherwise the procedure is similar to that of Experiment 1A.

#### Method

Participants. Eleven paid volunteers or introductory psychology students participated in this experiment for payment or credit. One participant was excluded for poor performance (<70% accuracy).

Design. The experimental design was similar to that used in Experiment 1 except for the following changes. Each participant was exposed to only one invariant sequence with eight distractor items. A different invariant sequence

was used for each participant. This sequence was preceded and followed by variable-length front and back ends that was seven items in total length, displacing absolute target position from the onset of the trial. This procedure allowed us to randomly vary absolute time from trial onset to target onset. Thus any learning observed in this experiment cannot be due to a stop-watch type learning mechanism. As before, the order of letter identities was randomized on every trial.

On every trial in the Practice, Training and Testing Phases, participants searched for a predefined target, either an X or K. A target was present on every trial. The experiment began with a 16-trial practice session. This was followed by 12, 16-trial blocks of Training Phase. This was followed by the Testing Phase. There were two blocks of each Testing Phase (four blocks total). The Testing Phase was followed by a Recognition Phase, consisting of 16 interleaved trials, eight variations of the invariant sequence used in the Training Phase and eight novel sequences.

Stimulus Materials- Training and Testing Phases. The sequence was composed of the following eight durations: 52, 104, 200, 333, 650, 910, 1105 and 1430 ms, each appearing once per sequence. The variable-ends were newly generated on each trial, and employed a random sample of the same durations used in the invariant sequence. For instance, the first occurrence of sequence 1 could have variable-end durations of 104 and 52 ms, and the second occurrence of sequence 1 could have durations of 333 and 200 ms. The target always occurred in the ninth position of the invariant sequence and always appeared for 80 ms. The total

Temporal Contextual Cueing sequence length was 16 letters. The same 16 alphabetic letters in variable random order were used on every trial.

The Testing Phase contained two interleaved conditions. Test 1 presented a novel sequence and the absolute timing of the sequence was equal to the absolute timing of the sequence in the Training Phase because it used the same number of durations, and the same set of durations.

Test 2 assessed the specificity of cueing from the invariant sequence. Test 2 used the same invariant sequence as in the Training Phase but target placement in the invariant sequence was manipulated. Either a single letter or four letters of randomly chosen durations were inserted between the invariant sequence and the target, breaking up the temporal configuration. The duration of each of the extra elements was randomly chosen from the list of durations used in the invariant sequences.

Stimulus Materials- Recognition Phase. Participants were exposed to eight novel sequences and eight variations of the invariant sequence. Each variation had a different variable front end and back end. The front and back ends were randomly generated and differed from each other and also from those seen during training.

Procedure. Participants performed Practice, Training, and Testing Phases, as in Experiment 1A. RT and accuracy were recorded and feedback was given. At the end of the experiment, participants completed a forced-choice recognition test in which no feedback was given. Participants were informed that half of all sequences they would see were old and half were new. The Recognition Phase was similar to the test used in Experiment 1A, except

that participants were required to make two responses. First, participants located the target in the sequence and made a speeded response. After making this response, participants made a decision as to whether or not they recognized the sequence. This design was used because the best way to test for explicit awareness is to closely match the testing procedure to the training procedure (Shanks & St. John, 1994). RT and accuracy were recorded for the first response, and only accuracy was recorded for the second response.

### Results

Blocks of two were grouped into epochs, creating six training epochs and one testing epoch. Epochs were smaller than in Experiment 1A to equate the number of trials in each testing condition to the number of trials in each training epoch. Mean correct performance was 91%. Accuracy by epoch data are reported in Table 1. Participants became more accurate over time,  $F(6,54) = 2.97$ ,  $p < .02$ , (including the Test 1 epoch),  $F(6,54) = 2.80$ ,  $p < .02$  (including the Test 2 epoch). This effect was primarily due to an increase in accuracy between the first and second epochs of training ( $t(9) = 2.8$ ,  $p < .007$ ) possibly due to enhanced familiarity with the targets or key mappings. There was no significant difference in accuracy between the two testing conditions,  $F < 1$ .

RT became faster over the course of training, 1:  $F(5, 45) = 3.36$ ,  $p < .012$  reflecting a general learning effect (Figure 3). t-tests show that RTs became slower in both Testing Phases, when compared to baseline performance in epoch 6:  $t(9) = 2.54$ ,  $p < .014$ ; and  $t(9) = 2.49$ ,  $p < .016$ , for Test 1 and Test 2, respectively. The decrease in Test 1 performance suggests that participants were indeed learning the invariant

Temporal Contextual Cueing sequence, and not an absolute overall time duration. Poorer performance in Test 2 suggests that cueing was highly specific to the learned onset of the target. Performance was significantly impaired when the target appeared at the wrong time with respect to the invariant sequence.

Explicit Recognition Test. Mean accuracy was 50%. Mean hit rate was 26% and mean false positive rate was 28% ( $F < 1$ ).

### Discussion

These results show that when absolute duration was controlled, learning of duration sequences was based on the ordering of elements in the sequence. This suggests that the learning seen in Experiment 1A was largely due to encoding of the sequence durations, not to a stop-watch type learning mechanism in which absolute (or additive) sequence durations were learned. The learning in Experiment 1A may have been partially due to an effect of absolute duration, but it may be a less important component than the serial order of durations or the rhythm of the sequences.

An associative learning mechanism may account for the learning demonstrated in these data. When the temporal association between the invariant sequence and the target was broken in Test 2 by inserting extra elements between target and invariant sequence, each taking up additional time in the duration sequence, RTs became slower. This result can be explained by a non-associative account also; the main conclusion is that temporal contextual cueing does not reflect a benefit from a general, vague expectation that something important was impending. Because Test 1 demonstrated that the

learning was not of absolute time from trial onset to target, we can assume that the change in absolute time created by the manipulation in Test 2 was not the cause of the disruption in performance. Rather, the disruption is most likely due to the broken temporal configuration. Similarly, Olson and Chun (in press) found that insertion of irrelevant material between a target and a repeated spatial configuration disrupted spatial contextual cueing. In addition, contextual cueing is disrupted when the target appears in variable locations across repetitions of spatial context (Chun & Jiang, 1998).

Further tests of the learning mechanism should be conducted in the future. For instance, the target could be randomly moved about, relative to the invariant sequence, during the training phase. Based on the results of Chun and Jiang's (1998) spatial contextual cueing studies, we predict that there should be little temporal cueing in such instances.

As before, explicit recognition was poor. This may be due to the possibility that participants focused more on the identity of the letters that carried no invariant, sequentially repeated information. The results indicate that observers can learn invariant sequential rhythm information without an explicit, conscious effort to do so. Such learning benefited search performance.

## **Section 2. Temporal contextual cueing from invariant identity sequences**

### **Experiment 2**

In most everyday situations the identity of visual information may be more important than the duration. Oftentimes we experience one thing happening after another where the identity of one thing predicts *what* will

Temporal Contextual Cueing come next. For instance, when we pass a certain set of houses on the way to work, the ordered sequence of neighborhood images may cue us that a stop sign is rapidly approaching. An invariant ordering of event identities also forms an important type of temporal structure and context that may be learned to improve performance (Jiménez & Méndez, 1999).

We examined whether or not implicit learning of a sequence of letter identities would cue attention to an upcoming important event, e.g. a target. The paradigm was similar to that used in Section 1 except that the sequence of letter identities was invariant, and the sequence of durations was random. The invariant sequence always began on the first frame and continued until the target appeared. Each invariant sequence was correlated with the onset of a target, either X or K, immediately following the invariant sequence. The question is whether that learning of letter identity sequences could cue attention to a target's sequential position.

### Method

Participants. Ten paid volunteers or introductory psychology students participated in this experiment for payment or credit.

Design and Procedure. Participants searched for a predefined target, either an X or K, that appeared on every trial. Participants completed eight trials of Practice Phase, 20 eight-trial blocks of Training Phase, and four eight-trial blocks of Testing Phase. At the end of the experiment, participants completed 16 trials of a Recognition Phase.

The paradigm is similar to that of Experiment IA, except that in this experiment, the sequential order of the

letters, and not durations, was manipulated. Figure 4 illustrates the stimulus design. Eight invariant letter sequences, from 3 to 10 letters in length, were newly generated for each participant. Each invariant sequence was shown one time per block. The invariant sequence always began on the first frame and continued until the target appeared. After each invariant letter sequence, a randomly chosen target (e.g. X or K) appeared. Thus each invariant letter sequence predicted when a target would occur, but did not predict what the target would be. The target letter was presented for 80 ms. Each invariant sequence was followed by a tail-end of randomized sequence that equated overall sequence length to 15 letters. Presentation of sequences was randomized. Letter stimuli and event durations were identical to those of Experiment 1A. However, the particular durations were randomized for each trial.

The Testing Phase was also similar to that of Experiment 1A. Participants were exposed to sequences that were identical to the sequences in the Training Phase except that the order of the letters was randomized. Thus letter identities were preserved between training and testing but the sequential order of the letters was not. Testing Phase onset was invisible to participants. To match sequences to the Training Phase, all invariant sequences were 3 to 10 items long with targets in the fourth through eleventh positions, respectively. This was followed by a random tail-end sequence, equating sequence length at 15 letters. The same event durations and letters were used for trial and test sequences.

The recognition test was similar to previously discussed tests, except for a slight difference in the instructional

Temporal Contextual Cueing wording. Specifically, participants were told " You may not have noticed, but there was a sequential pattern to some of the letter sequences you saw. For most of the trials, the letters followed a particular temporal order, such as in the alphabet (a,b,c,d,...). If you noticed this order, you could have used it to predict what the target was going to be. In this next section we want to see if you recognize any of the sequences."

The entire session took 40 minutes.

### Results

Four blocks were grouped to form an epoch. There were 6 epochs. Mean correct performance was 86%. Accuracy by epoch is reported in Table 1. Participants became significantly more accurate over time,  $F(5, 45) = 4.5$ ,  $p < .002$ . This improvement appears to be driven by general motor or skill learning, and not learning of specific sequences, since accuracy was also high in the Testing Phase.

In the Training Phase, a repeated measures ANOVA of RT found a non-significant effect of epoch,  $F(4, 36) = 1.46$ , *n.s.* To look at sequence-specific learning, a t-test was conducted between the last epoch of the learning condition and the testing condition. This analysis found that participants became slower in the Testing Phase,  $t(9) = 2.66$ ,  $p < .026$ . Results are shown in Figure 5.

Explicit Recognition Test. Mean accuracy was 49%. Mean hit rate was 32% and mean false positive rate was 34%. No statistically significant differences were found between conditions ( $F < 1$ , *n.s.*).

### Discussion

This experiment shows that identity of sequentially organized items can cue attention to a target's appearance

and identity in a sequence. Previous results (Chun & Jiang, 1999) showed that the identity of spatially arrayed distractors can cue attention to a correlated target shape. In their task, participants searched for a vertically symmetrical novel object. Implicit learning was manipulated by correlating particular target identities with the particular identities of groups of distractors. Locations were randomized. They found that invariant distractor identities cued a correlated target identity. A subsequent recognition test showed that learning was implicit. The present results extend Chun and Jiang's results to the time domain. Invariant identity sequences provided a temporal context whereby target onset and identity could be predicted by the preceding events. These results suggest that the temporal context, like spatial layout (Chun & Jiang, 1998) or spatial identity (Chun & Jiang, 1999) context, can provide a rich source of predictive information (Jiménez & Méndez, 1999). It is possible that invariant information of this sort is used in one of our most important visual behaviors, navigation. Although navigation can occur in the absence of landmarks by dead-reckoning, landmarks are well-known to be an easily accessible and commonly used navigation tool (Gallistel, 1990).

The results of Experiment 2 were weaker than that reported in the pilot experiment and also in the experiments of Section 1. It may be the case that sequences of letter identities are easily associated with a target letter identity and response but an association between sequence letter identities and target sequential position is more difficult to form. Frensch (1998) suggested that certain stimulus relationships form easily under implicit conditions while others

Temporal Contextual Cueing are nearly impossible to form. Although our data do not allow a direct test of this claim, it seems that sequences of letter identities are more effective in predicting *what* target letter may occur than in predicting *when* a target event may occur.

Participants were unaware that they were learning or being exposed to invariant letter sequences, as shown by the recognition data.

### **Section 3. Temporal contextual cueing from invariant spatio-temporal sequences.**

#### **Experiment 3A**

When we look around, most visual events are defined over both space and time. As we navigate through space, the visual scene is constantly changing and updating our representation of where we are in space and time relative to the spatio-temporal context of other events. The experiments in this section examine how these two variables, space and time, interact to guide attention. The previous experiments demonstrate a response time benefit for targets preceded by a sequentially ordered sequence of letters or event durations. In this experiment we examined how invariant spatio-temporal sequences can cue attention to a location in space.

To explore these issues a modified version of the previously described task was used. As before, the experiment had Training, Testing, and Recognition Phases. During the Training Phase we used a sequential visual-search task in which participants searched for a predefined target. The stimuli were presented in a rapid constant-duration stream. Each frame was either a distractor frame containing two distractors, or a target frame

containing one target and seven distractors.

Each trial started with random frames, followed by an invariant spatio-temporal sequence, and then the target frame. Last, there was a tail-end of randomly assigned frames. Each frame could differ in the locations of the distractors appearing on them, so a fixed series of these distractor frames defined a spatio-temporal sequence. Invariant spatio-temporal sequences were defined by a sequence of distractor frames, repeated throughout the experiment (See Figure 6). In other words, the distractor locations changed in an invariant manner preceding the target frame. Eight spatio-temporal sequences were created for each participant. Thus the temporal sequence in this experiment, and the other experiments discussed in Section 3, resided in the sequential order of a small number of locations. Because duration was not varied, the temporal sequences had an isochronous rhythm. Each spatio-temporal sequence correlated with the spatio-temporal position of a randomly chosen target. To vary absolute target placement relative to the trial onset, a random "front-end" sequence was placed in front of each sequence at the beginning of each trial, obscuring the beginning of the invariant sequence. The stimulus design is illustrated in Figure 6.

The participants' task was to look for the target in the stream of distractors. The prediction was that participants would encode the invariant sequences and use this information to guide attention to the location of the target on the target frame. This would be shown by faster responses in the Training Phase.

During the Testing Phase, participants did the same task.

### Temporal Contextual Cueing

Unknown to them, the sequences had been changed so that they no longer contained invariant information, that is, each sequence contained a random ordering of the same distractor locations as used in the Training Phase. We expected this to impair performance. This allowed us to compare performance at the end of training to performance in the testing condition thereby assessing sequence-specific learning.

After testing, participants took a recognition test in which they viewed each sequence and responded as to whether or not they recognized the spatio-temporal patterns.

#### Method

Participants. Eleven paid volunteers or introductory psychology students participated in this experiment for payment or credit. One participant was excluded for poor performance (<60% in any epoch).

Design. The temporal context was provided by the sequence of distractor locations appearing prior to the target frame. Eight spatio-temporal invariant sequences were newly created for each participant and were used for the duration of the experiment. Each sequence appeared once per block. Participants searched for a predefined target, present on every trial. The target frame always immediately followed the invariant sequence. Target identity was randomized for each trial, and hence did not correlate with any of the invariant sequences. All frames in all sequences were presented for the same amount of time, hence, the invariant patterns had isochronous rhythms. One practice block of eight trials was followed by 18 Training Phase blocks (8 trials per block), and six Testing Phase blocks (8 trials per block). Lastly, 16 trials of Recognition Phase were presented.

Stimuli. Each trial consisted of 16 frames (one target frame and 15 distractor frames), sequentially presented in a rapid stream. Each invariant spatio-temporal sequence was seven frames long with the target appearing on the eighth frame. Absolute temporal position of the target was varied by modulating the length of random (“junk”) front and back end sequences. The front end sequence was either 3, 4, or 5 frames long and was matched to a backend sequence of 5, 4, or 3, or frames. The eight invariant spatio-temporal sequences were randomly paired with the junk front and back ends at the beginning of each block.

All stimuli were superimposed on a faint white dot grid, size 200 x 200 pixels, on a gray background. The grid was made up of an inner and outer circle, each containing eight positions. There were two green distractors, an offset 'L' and a Christian cross.frame Distractor items always appeared along the ends of an invisible diagonal equidistant from central fixation. This reduced the number of potential locations from 16 to eight. However, this technique, combined with a small font size (approximately .69 cm) served to discourage eye movements. Each pair of locations was used twice per sequence. The target frame contained one target, a “T” rotated to the left or right, and one of the distractors. Target identity was randomly chosen at the beginning of each trial.

In the Testing Phase the task was identical to that of the Training Phases except that the sequential order of the distractor elements making up the spatio-temporal sequence was randomized. Targets and distractors were the same colors, shapes, and appeared in the same locations as in the

Temporal Contextual Cueing Training Phase. Also, targets appeared in the same serial positions as in the Training Phase to control for frequency effects of target serial position. Again, each location pair was used twice per sequence. Testing Phase onset was invisible to participants.

The Recognition Phase consisted of eight invariant sequence trials interleaved with eight novel sequence trials.

Procedure. The experiment was carried out in normal room illumination held constant for all participants. The experiment was self-paced. The participant began each trial by pressing the space bar on the computer keyboard. A fixation dot appeared for 500 ms and then went off for 13 ms. Participants were asked to minimize eye movements by fixating on the center. The stream of stimuli then appeared successively without interstimulus blanks at the same location for durations of 266 ms each. The sequence was followed by a fixation dot. The participants’ task was to monitor the stream and report the direction of the bottom of the target letter , entering their response as either “z” for left or “/” for right. Responses were entered as quickly as possible into the computer using the keyboard. Participants were encouraged to enter their response at the earliest time possible even if the trial was not yet complete. Positive feedback was given for all RT trials in the form of high-pitched beeps, negative feedback in the form of a low-pitched buzz.

The Testing Phase was followed by the Recognition Phase. Participants were informed about the purpose and manipulations in the experiment and were then asked to passively view each of the eight sequences plus eight novel sequences randomly intermixed. After

viewing each sequence the participant was asked by the computer whether or not they recognized the sequence. They responded by pressing a "y" key for yes, or a "n" key for no. No feedback was given.

### Results

Block data were grouped into four epochs, each containing six blocks. Mean correct performance was 81%. Accuracy by epoch data is reported in Table 1. The main effect of accuracy approached significance,  $F(3,27) = 2.708$ ,  $p < .065$ , suggesting that participants became more accurate over time. However there was no significant difference between the last epoch of the learning condition and the Testing Phase epoch ( $t(9) = .726$ ,  $p > .474$ ), suggesting that the learning effect revealed by accuracy was not sequence-specific.

In the repeated measures ANOVA of Training Phase RTs, there was no main effect of epoch,  $F(3,27) = 1.87$ ,  $p < .158$ , reflecting the shallow learning curve. There was a significant effect of condition between the last epoch of the learning condition and the Testing Phase epoch,  $t(9) = 2.218$ ,  $p < .035$ , as shown in Figure 7. Thus, participants encoded the spatio-temporal structure of the events preceding the target and used this information to facilitate detection of the target.

Explicit Recognition Test. Mean accuracy was 55%. Mean hit rate was 58% and mean false-positive rate was 48%, ( $F < 1$ ).

### Discussion

The data from this experiment show that invariant spatio-temporal sequences are implicitly learned and can make search performance faster and more accurate. This manipulation linked

### Temporal Contextual Cueing

a spatio-temporal sequence of events to a target location at a specific time, suggesting that visual sequence learning can facilitate attentional deployment to a location in space. Recognition performance was poor, reflecting the implicit nature of this learning.

Our data do not allow us to determine whether the spatio-temporal sequences were cueing the onset of the target and/or the location of the target. We conjecture that both temporal and spatial attributes were cued. Currently, existing data from our lab suggest that at least target location can be cued by the invariant spatio-temporal sequences. This was shown in an experiment that presented additional cues for the temporal onset of the target. Significant contextual cueing was obtained, and this benefit was likely to have been driven by cueing of the target location because visual cues already existed in all conditions to facilitate detection of target onset. It will be interesting to separate the role of spatial cueing and temporal cueing more systematically in future work.

Another important question concerns is what aspect of the temporal sequence constitutes the cue for the target's temporal and spatial location? Participants could be using the entire temporal sequence or only part of the sequence to cue their attention. It is possible that frames that are closer to the target are more salient and provide greater cueing than frames that are at the beginning of each invariant sequence. Experiments 3B and 3C examine this question.

### **Experiment 3B**

We designed an experiment that was similar to Experiment 3A, except that there were three conditions in the Testing Phase. The first condition was

the “New” condition, where all elements were randomly selected, except the target locations. The second condition was the “Old” condition, where all elements were invariant and identical to the sequences seen during the Training Phase. The third condition had all elements randomly chosen except for the target and the item immediately preceding the target (the target-1 element). Thus, this condition allowed us to examine whether or not cueing would occur from just a single frame instead of a more global temporal sequence. To increase statistical power, participants were exposed to 16 invariant sequences instead of 8.

#### Method

Participants. Seventeen paid volunteers participated in this experiment for payment or credit. One participant was excluded for poor performance (<60% in any epoch).

Design and Procedure. The design was similar to Experiment 3A, except where noted. As before, distractor items always appeared on an invisible line balanced around the center of the circular grids. There were two white distractors, offset 'L's. Distractors varied from frame to frame and from trial to trial. Item font size was approximately .69 cm. The target frame contained one target, a rotated “T”, and one of the distractors. Target identity was randomly chosen at the beginning of each trial. Unlike Experiment 3A, the distractors were not allowed to occur in the same location consecutively. Also, the sequences were shortened so that each invariant sequence was five items long (excluding the target), and total sequence length was ten items. Each stimulus frame appeared for 240 ms, without interstimulus blanks.

#### Temporal Contextual Cueing

Each block contained sixteen trials. One practice block (16 trials) was followed by eight Training Phase blocks. This was followed by four Testing Phase blocks, each containing 48 trials, and one Recognition Phase block (32 trials).

#### Results

Block data was grouped into three epochs, each containing four blocks. Mean correct performance was 95%. Accuracy by epoch data is reported in Table 1.

Training Phase. Learning was most apparent in the accuracy measure. Accuracy became significantly higher during the Training Phase,  $t(15) = 2.845$ ,  $p < .008$ . RT did not become significantly faster,  $F(1,15) = 2.38$ ,  $p > .144$ , reflecting the shallow learning curve (12 ms faster in epoch 2).

Testing Phase. RTs from all three test conditions were submitted to a repeated measures ANOVA. There was a main effect of condition  $F(2,30) = 3.99$ ,  $p < .029$ . t-tests between each condition show that both the old and the t-1 old conditions were significantly faster than the new condition ( $t(15) = 2.19$ ,  $p < .036$ , and  $t(15) = 2.64$ ,  $p < .013$ , respectively). There was no difference between the old and t-1 old conditions,  $t(15) = .446$ , *n.s.* This suggests that the t-1 frame may have provided much of the cueing benefit. These results are shown in Figure 8.

Explicit Recognition Test. Mean accuracy was 44%. Hits averaged 46% and false-positives averaged 58%. There was no statistically significant difference ( $F < 1$ ).

#### Discussion

These results suggest that the t-1 element in an invariant spatio-temporal sequence occupies a particularly salient position. It is possible that there is a salience gradient around the target

location such that items which are further away provide less cueing than closer elements. However, it is unlikely that the t-1 element provides all of the cueing benefit in the studies presented in this paper. In Experiment 3B, each location in the invariant sequence occurred once, thus the t-1 element provided unique information about the target. However, this was not true in previous experiments. In Experiment 1A, most of the invariant sequences contained durations that repeated. For instance, a 10-element invariant sequence would have had two repetitions of each duration, thus the t-1 element could not have provided unique cueing information. Nevertheless, to study this possibility in more detail for spatio-temporal sequence learning, we conducted one additional experiment.

### Experiment 3C

Experiment 3C tested whether participants could use longer-range temporal context to cue attention. To test this, we designed an experiment similar to the last, except that the frame preceding the target, the target-1 item, was the same on every trial. Thus this frame provided no sequence-specific information about the onset and location of the target.

#### Method

Participants. Eight students participated in this experiment for payment.

Design and Procedure. The stimuli were similar to those used in Experiment 3B. Each sequence contained a five-item long invariant sequence, a target, one fixed element in the t-1 position, and five junk elements (total length=12 frames). Each frame was presented for 187 ms with no interstimulus interval. The presentation

Temporal Contextual Cueing rate was increased to lower accuracy performance below ceiling.

The distractors immediately preceding the target frame always appeared in the same locations, in training, testing, and recognition. Two linearly opposite positions on the grid were reserved for the t-1 frame. None of the other elements in the invariant sequence were allowed to occupy this position. However, front-end and back-end junk elements were allowed to appear in this position. In other respects, the Training Phase was similar to that used previously.

The onset of the Testing Phase was invisible to participants. The Testing Phase contained two conditions, Old (invariant sequences) and New (randomly generated sequences). The two conditions were randomly interleaved.

Each block contained seven trials. One practice block (14 trials) was followed by 20 training blocks (seven trials each). Every two blocks was separated by a break. This was followed by four testing-phase blocks, each containing 14 trials, with a break after each block, and one recognition test block (14 trials).

The experiment lasted 30 minutes.

#### Results

Mean correct performance was 89%. Block data from the Old condition were grouped into six epochs, each containing four blocks. Accuracy by epoch data is reported in Table 1. The main effect of accuracy was significant in the Old condition,  $F(5,35) = 3.59$ ,  $p < .016$ , suggesting that participants became more accurate over time. However, there was no significant difference in accuracy between the New and Old conditions in the Testing Phase ( $F < 1$ ),

suggesting that the accuracy benefit was due to general practice effects.

RTs became faster over time in the Old condition, shown in the repeated measures ANOVA,  $F(5,35) = 4.83$ ,  $p < .0018$ . A separate ANOVA compared the Old and New conditions in the Testing Phase and found that the Old condition was faster than the New,  $F(1,7) = 8.51$ ,  $p < .023$ . Thus, participants were able to use invariant spatio-temporal information that occurred earlier than the t-1 position. These results indicate that the t-1 element is not the only source of temporal cueing. Results for this analysis are shown in Figure 9.

Explicit Recognition Test. Due to computer failure, data were not recorded for one subject. For the remaining seven subjects, mean correct was 55%, hits averaged 46% and false-positives averaged 36%. There was no statistically significant difference ( $F < 1$ ).

#### Discussion

The results from Experiment 3C show that participants can successfully encode invariant global temporal sequences. Because the t-1 element was the same in every sequence, there was no predictive advantage for Old versus New conditions. Thus any RT differences between the Old and New conditions had to be due to encoding of earlier sequence information. As before, learning was implicit.

#### **General Discussion**

The temporal structure of visual events can be used to predict when something important is going to occur, what it might be, and where it may occur. In Section 1, we found that invariant sequences of event durations can cue attention to a point in time.

#### Temporal Contextual Cueing

Invariant sequences led to faster target identification latency. Temporal sequence learning was sequence-specific. Specificity in serial order between the sequence and target position appears to be critical.

The experiment in Section 2 manipulated event identity. We found that invariant sequences of letter identities can guide attention to a serial position and also cue the identity of an upcoming event.

Experiments in Section 3 used more complex visual displays that presented an invariant sequence of events defined over both space and time. There was a RT benefit for invariant spatio-temporal sequences compared to randomized sequences containing the same low-level information. An additional experiment suggested that much of the cueing benefit was derived from the element immediately preceding the target frame (e.g. the t-1 element). However, we confirmed in Experiment 3C that earlier items in the invariant sequence can also cue the target's appearance and location. Learning was found to be implicit in all cases, as assessed by a forced choice recognition task. These findings suggest a model of sequential learning where items near the target have stronger cueing power, because of their salient position. However, other elements in the sequence are also encoded to cue attention.

In summary, we introduced a new paradigm termed temporal contextual cueing, and found that (1) temporal structure in perception can guide attention to a serial position, a point in space, or can cue the identity of an upcoming event; (2) learning of stimulus sequences can occur without intention or awareness; (3) this type of learning is sequence specific and may be

associative; and (5) the element immediately preceding the target is particularly salient and may provide more cueing power than other elements in the sequence.

Previous researchers have also proposed that dynamic events can direct attention. One particularly well-articulated model postulates that attention becomes “entrained” by the rhythmic structure of stimuli (e.g. Jones & Yee, 1997; Large & Jones, 1999). Evidence for this theory is found in auditory perception studies. We emphasize a different time-based variable in this paper: sequential order. Our data confirm a more general version of the attentional entrainment theory: that many types of dynamic structures can guide visual attention to specific temporal positions, target identities, or spatial locations.

Temporal associations may be important for recognition (Wallis & Bulthoff, 1999; Wallis & Rolls, 1997). The results from Section 2 are pertinent to this idea. These results suggest that serially ordered event sequences are easily learned and can be used to predict when an important event will happen in the future. This type of learning can be implicit.

These experiments show that the temporal structure of the visual world contains information that can assist visual processing, just as invariant spatial structure does (Chun, 2000; Chun & Jiang, 1998). Recently Coull and Nobre (1999) developed a temporal cueing task analogous to the Posner spatial cueing task and demonstrated a behavioral advantage for explicit temporal cueing (see also Kingstone, 1992). Our results show an advantage for implicit temporal cueing. Our task is more ecological than standard cueing

Temporal Contextual Cueing tasks in the literature because the implicit nature of the “cue”, that is, the invariant sequences, mimics natural experiences in vision more closely. We are rarely given punctate, explicit instruction when something will occur. Rather, we can easily anticipate when something is going to happen by implicitly encoding an ongoing, continuous temporal context. Most of visual experience is structured and invariant so it makes sense that the brain encodes the invariant information to facilitate perception when those temporal structures are re-experienced.

Not only is temporal structure encoded, it works in concert with predictive spatial information. Section 3 demonstrated how spatio-temporal structure can cue attention to target onsets and locations. In a prior study, spatio-temporal context defined by motion trajectories allowed participants to implicitly predict where a target would occur (Chun and Jiang, 1999). Like Chun and Jiang’s results, the results presented in Section 3 suggest that spatial and temporal cues are integrated to facilitate perceptual performance. In fact, a number of findings converge to reveal a tight, interactive relationship between spatial attention and temporal attention (Coull & Nobre, 1999; Husain, Shapiro, Martin, and Kennard, 1997; Jiang & Chun, in press; Marois, Chun, & Gore, 2000).

Implicit sequence learning has previously been studied using the SRT task (Lewicki et. al., 1988; Mayr, 1996; Nissen & Bullemer 1987; Willingham, Nissen, & Bullemer, 1989). Significant differences exist between the temporal contextual cueing task and the SRT task. First, unlike temporal contextual cueing, in the SRT, each stimulus in a sequence is associated with a response. Second,

## Temporal Contextual Cueing

SRT invariant sequences of stimuli are typically correlated with invariant sequences of motor responses. Studies that have tried to disentangle stimulus learning from response learning (Hoffmann & Sebold, 1996; Keele et al., 1995) suggest that the bulk of the learning seen in classic SRT manipulations is due to motor skill acquisition, while other studies suggest that much of the learning is stimulus driven (Cohen, Ivry, & Keele 1990; Mayr, 1996). Data from brain imaging studies of the SRT task favor the former conclusion (Grafton, Hazeltine, & Ivry, 1995; Hazeltine, Grafton, & Ivry, 1997; Hazeltine, Helmuth, & Ivry, 1997) since learning-related activity is most consistently seen in motor areas, though there does appear to be a perceptual component in some cases (see Curran, 1997 for a review). In conclusion, our paradigm may be more advantageous for studying effects of temporal structure in perception, while the SRT task is well-suited for studying motor sequence learning. It should be noted that explicit learning may have been attenuated in some of our experiments, due to the fact that recognition was probed after inclusion of multiple Test Phase blocks with random sequences. However, Experiments 3B and 3C used a different Test Phase design in which invariant and random sequences were interleaved. Since the recognition results were similar in these experiments to those reported in the other experiments, it seems likely that the inclusion of Test Phase blocks prior to recognition cannot account for the lack of explicit recognition.

The importance of temporal structure has received less attention from vision researchers than sister disciplines such as neuroscience and animal

learning which have long recognized the significance of serial order for behavior (e.g., Lashley, 1951). Disorders in temporal processing are found in certain neurological populations and may have an impact on some types of visual processing. For instance, patients with unilateral visual neglect are unaware of people or objects appearing in the visual hemifield contralateral to their lesion. On the surface this appears to be a purely spatial deficit yet Husain et al(1997) showed that they also have a disorder in directing attention in time as evidenced by the inability to process two stimuli if they occur within 1.5 seconds of each other (attentional blink: Broadbent & Broadbent, 1987; Chun & Potter, 1995; Raymond et al., 1992). Thus, information about spatial location and time of occurrence may be intimately linked in certain cognitive operations like attentional deployment. Other researchers have reported deficits in both estimating duration and performing sequential tasks in Parkinson's disease (Artieda, Pastor, Lacruz, Obeso, 1992; Jackson, Jackson, Harrison, Henderson, et al., 1995). Deficits in performing sequential tasks have also been reported in schizophrenia (Dominey & Georgieff, 1997; Green, Kern, Williams, McGurk, Kee, 1997). Outside the laboratory it is clear that such knowledge is critically important. For instance, in basketball, invariant information about the speed of other player's movements, the manner of movements and the final trajectory of movement greatly constrain a player's decisions about how, when, and where to move (e.g., Chun & Jiang, 1999). Time and space are the most obvious variables that we can use to structure information (Hoffman and Koch, 1998). As such it should be of no surprise that the brain

## Temporal Contextual Cueing

uses them, either singularly or in tandem, to detect and utilize invariant spatio-temporal structure in an information-overloaded perceptual world.

## Author Note

Ingrid R. Olson, Department of Psychology, Yale University, and Marvin M. Chun, Department of Psychology, Vanderbilt University.

The research was supported by National Science Foundation (NSF) Grant BCS-9817349. Ingrid R. Olson was supported by a predoctoral fellowship from NSF and doctoral grants from the American Psychological Association, Yale University, and the John B. Enders Foundation. Correspondence concerning this article should be addressed to Ingrid R. Olson, Department of Diagnostic Radiology, Yale School of Medicine, 333 Cedar Street, Fitkin Bldg, Rm F14, New Haven, CT 06510. Electronic mail should be sent to iolson@boreas.med.yale.edu.

## Acknowledgements

We would like to thank Yuhong Jiang for her help with analysis, and the dissertation committee of I. Olson, Frank Keil, Carol Fowler, Paul Bloom, and Truett Allison, for helpful discussion

## References

Artieda J., Pastor M.A., Lacruz F., Obeso J.A. (1992). Temporal discrimination is abnormal in Parkinson's disease. Brain, 115, 199-210.

Broadbent, D. E., Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. Perception & Psychophysics, 42, 105-113.

Brown, G.D.A., Hulme, C., Preece, T. (2000). Oscillator-based memory for serial order. Psychological Review, 107, 127-181.

Chomsky, N. (1957). Syntactic structures. The Hague: Mouton.

Chun, M. M. (2000). Contextual cueing of visual attention. Trends in Cognitive Sciences, 4, 170-178.

Chun, M. M., Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. Cognitive Psychology, 36, 28-71.

Chun, M. M., Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. Psychological Science, 10, 360-365.

Chun, M. M., Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance, 21, 109-127.

Chun, M. M., Wolfe, J.M. (2000). Visual Attention. In E. B. Goldstein (Ed.), Blackwell Handbook of Perception.

Clegg, B. A., DiGirolamo, G.J., Keele, S.W. (1998). Sequence learning. Trends in Cognitive Sciences, 2, 275-281.

Cleeremans, A., McClelland, J.L. (1991). Learning the structure of event sequences. Journal of Experimental Psychology: General, 120, 235-253.

Cohen, A., Ivry, R.I., Keele, S.W. (1990). Attention and structure in sequence learning. Journal of Experimental Psychology: Learning, Memory and Cognition, 16, 17-30.

Coull, J. T., Nobre, A.C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. The Journal of Neuroscience, 18, 7426-7435.

Curran, T. (1997). A cognitive neuroscience perspective. In M. Stadler, & P. Frensch, (Eds.), The Handbook of Implicit Learning. Thousand Oaks: Sage Publications.

Doeringer, J. A., Hogan, N. (1998). Serial processing in human movement production. Neural Networks, 11, 1345-1356.

Dominey, P. F. Georgieff, N. (1997). Schizophrenics learn surface but not abstract structure in a serial reaction time task. Neuroreport: an International Journal for the Rapid Communication of Research in Neuroscience, 8, 2877-2882.

Elman, J.L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. Machine Learning, 7, 195-224.

Eriksen, C.W., Hoffman, J.E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. Perception & Psychophysics, 12, 201-204.

Estes, W.K. (1985). Memory for temporal information. In J.A. Michon & J.L. Jackson (Eds), Time, Mind, and Behavior. Berlin, Germany: Springer-Verlag.

Frensch, P. (1998). One concept, multiple meanings. In M. Stadler & P. Frensch (Ed.), The Handbook of Implicit Learning. Thousand Oaks, CA: Sage.

## Temporal Contextual Cueing

Gallistel, C.R. (1990) The Organization of Learning. Cambridge, MA: MIT Press.

Grafton, S. T., Hazeltine, E., Ivry, R. (1995). Functional mapping of sequence learning in normal humans. Journal of Cognitive Neuroscience, 7, 497-510.

Green, M.F., Kern, R.S., Williams, O., McGurk, S., Kee, K. (1997). Procedural learning in schizophrenia: Evidence from serial reaction time. Cognitive Neuropsychiatry, 2, 123-134.

Harman, K. L., & Humphrey, G. K. (1999). Encoding 'regular' and 'random' sequences of views of novel three-dimensional objects. Perception, 28, 601-615.

Hazeltine, E., Helmuth, L.L., Ivry, R.B. (1997). Neural mechanisms of timing. Trends in Cognitive Science, 1, 163-169.

Hazeltine, E., Grafton, S.T., Ivry, R. (1997). Attention and stimulus characteristics determined the locus of motor sequence encoding: A PET study. Brain, 120, 123-140.

Hoffmann, J., Sebald, A. (1996). Reiz- und reaktionsmuster in seriellen wahlreaktionen. Zeitschrift für Experimentelle Psychologie, 43, 40-68.

Hoffmann, J., Koch, I. (1998). Implicit learning of loosely defined structures. In M. Stadler, & Frensch, P. (Ed.), The Handbook of Implicit Learning. Thousand Oaks: Sage Publications.

Hunt, S. M. J. (1994). MacProbe: A Macintosh-based experimenter's workstation for the cognitive sciences. Behavior Research Methods, Instruments, & Computers, 26, 345-351.

Husain, M., Shapiro, K., Martin, J., Kennard, C. (1997). Abnormal temporal dynamics of visual attention in

spatial neglect patients. Nature, 385, 154-156.

Jackson, G.M., Jackson, S.R., Harrison, J., Henderson, L., et al. (1995). Serial reaction time learning and Parkinson's disease: Evidence for a procedural learning deficit. Neuropsychologia, 33, 577-593.

Jiang, Y., & Chun, M. M. (in press). The influence of temporal selection on spatial selection and distractor interference: An attentional blink study. Journal of Experimental Psychology: Human Perception & Performance.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? Journal of Experimental Psychology: Learning, Memory, & Cognition, 25, 236-259.

Jones, M.R., Yee, W. (1997). Sensitivity to time change: the role of context and skill. Journal of Experimental Psychology: Human Perception and Performance, 23, 683-709.

Jordan, M. (1995). The organization of action sequences: Evidence from a learning task. Journal of Motor Behavior, 27, 179-192.

Jordan, M.I. (1997). Serial order: A parallel distributed processing approach. In J.W. Donahoe, & V. P. Dorsel et al. (Ed), Neural-network models of cognition: Biobehavioral foundations. Advances in psychology, 121. Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Keele, S. W., Jennings, P., Jones, S., Caulton, D., et al. (1995). On the modularity of sequence representation. Journal of Motor Behavior, 27, 17-30.

Kingstone, A. (1992). Combining expectancies. Quarterly Journal of

Temporal Contextual Cueing Experimental Psychology, 44 [A], 69-104.

Klein, J. M., Jones, M.R. (1996). Effects of attentional set and rhythmic complexity in attending. Perception & Psychophysics, 58, 34-46.

Large, E. W., Jones, M.R. (1999). The dynamics of attending: How we track time-varying events. Psychological Review, 106, 119-159.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), Cerebral Mechanisms in Behavior. New York: John Wiley.

Lewandowsky, S., Li, S.-C. (1994). Memory for serial order revisited. Psychological Review, 101, 539-543.

Lewicki, P., Czyzewska, M., Hoffman, H. (1987). Unconscious acquisition of complex procedural knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 523-530.

Marois, R., Chun, M. M., & Gore, J. C. (2000). Neural correlates of the attentional blink. Neuron, 28, 299-308.

Mayr, U. (1996). Spatial attention and implicit sequence learning: evidence for independent learning of spatial and nonspatial sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 350-364.

Nissen, M. J., Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. Cognitive Psychology, 19, 1-32.

Perruchet, P., Amorim, M.-A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 785-800.

Posner, M. I., Cohen, Y. (1984). Components of visual orienting. In R. L. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Posner, M. I., Snyder, C.R.R., Davidson, B.J. (1980). Attention and the detection of signals. Journal of Experimental Psychology: General, 109, 160-174.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? Journal of Experimental Psychology: Human Perception & Performance, 18, 849-860.

Reber, A. S. (1989). Implicit learning and tacit knowledge. Journal of Experimental Psychology: General, 118, 219-235.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. Cognition, 70, 27-52.

Shanks, D. R., St. John, M.F. (1994). Characteristics of dissociable human learning systems. Behavioral and Brain Sciences, 17, 367-447.

Stadler, M. A. (1995). Role of attention in implicit learning. Journal of Experimental Psychology: Learning Memory and Cognition, 21, 674-685.

#### Temporal Contextual Cueing

Wallis, G., Rolls, E.T. (1997). A model of invariant object recognition in the visual system. Progress in Neurobiology, 51, 167-194.

Wallis, G., Baddeley, R. (1997). Optimal unsupervised learning in invariant object recognition. Neural Computation, 9, 883-894.

Wallis, G. (1998). Spatio-temporal influences at the neural level of object recognition. Network: Computational Neural Systems, 9, 265-278.

Wallis, G., Bulthoff, H. (1999). Learning to recognize objects. Trends in Cognitive Science, 3, 22-31.

Willingham, D. B., Nissen, M.J., Bullemer, P. (1989). On the development of procedural knowledge. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 1047-1060.

Willingham, D. B., Greenley, D.B., Bardona, A.M. (1993). Dissociation in a serial response time task using a recognition measure: comment on Perruchet and Amorim. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 1424-1430.

Willingham, D. B. (1999). Implicit motor sequence learning is not purely perceptual. Memory and Cognition, 27, 561-572.

Yantis, S., Jonides, J. (1996). Attentional capture by abrupt onsets: new perceptual objects or visual masking? Journal of Experimental Psychology: Human Perception & Performance, 22, 1505-1513.

## Table Captions

Table 1. Average accuracy by Epoch for all experiments. The last epoch is always from the Testing Phase, the preceding epochs are always from the Training Phase. Experiment 1B has two numbers in the Testing phase, denoting accuracy in Test 1 and Test 2, respectively.

Table 1

<i>Experiment</i>	<i>Epoch</i>						
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
1A	.85	.90	.88	.87			
1B	.85	.93	.92	.95	.93	.92	.88/.89
2	.80	.84	.88	.87	.88	.90	
3A	.76	.80	.85	.83			
3B	.92	.96	.96				
3C	.84	.90	.92	.92	.88	.89/.90	

## Figure Captions

Figure 1. A general schematic of the trials in Section 1. The differently spaced frames symbolize the varying event durations that made up the invariant sequences. The invariant sequence was immediately followed by the target frame, shown in shadowed font, for illustration. Note that the durations are held constant across repetitions, but the order of letters changes. Each invariant sequence was repeated one time per Training Phase block. During the Testing Phase, the invariant sequences were randomized.

Figure 2. The effect of invariant event duration sequences on target identification latency (response time measured from target onset) for Experiment 1A. The Training Phase of the Experiment 1A comprises the first 3 epochs. The Testing Phase, where randomized sequences were presented is shown in epoch 4. Error bars represent the standard error of the mean in this and all other figures.

Figure 3. The effect of invariant event duration sequences on target identification latency (response time measured from target onset) for Experiment 1B. The Training Phase of the experiment comprises the first 6 epochs. The Testing Phases, (Test 1: controlled randomized sequences (shown with solid line); Test 2: invariant sequence but with displaced target (shown with broken line)) are nearly overlapping in epoch 7.

Figure 4. A general schematic of the trials in Section 2. The invariant sequence was found in the order of letters preceding the target frame, shown in shadowed font, for illustration. The event durations were randomized across invariant sequence repetition. Each invariant sequence was repeated one time per Training Phase block. During the Testing Phase, the invariant sequences were randomized.

Figure 5. The effect of invariant event identity sequences on target identification latency (response time measured from target onset) for Experiment 2. The Training Phase comprises the first 4 epochs, the Testing Phase comprises the fifth.

Figure 6. A schematic of one invariant sequence found in a trial for Experiment 3A. The invariant sequence was found in the spatio-temporal ordering of the distractors in the frames immediately preceding the target frame. Each invariant sequence was repeated one time per Training Phase block. During the Testing Phase, the order of the distractor frames preceding the target frame followed no particular pattern.

Figure 7. The effect of invariant spatio-temporal sequences on target identification latency (response time measured from target onset) for Experiment 3A. The Training Phase comprises the first three epochs, testing the fourth. There is a statistically significant difference between the last epoch of the Training Phase and the Testing Phase.

Figure 8. The effect of invariant spatio-temporal sequences on target identification latency (response time measured from target onset) for Experiment 3B. Data are shown for the three conditions of the Testing Phase. There is a statistically significant difference between the new condition and both old conditions.

Figure 9. The effect of invariant spatio-temporal sequences, without a predictive t-1 element, on target identification latency (response time measured from target onset) for Experiment 3C. Data are shown for the Testing Phase (epoch 6). There is a statistically significant difference between the new and old conditions in the Testing Phase.

Figure 1

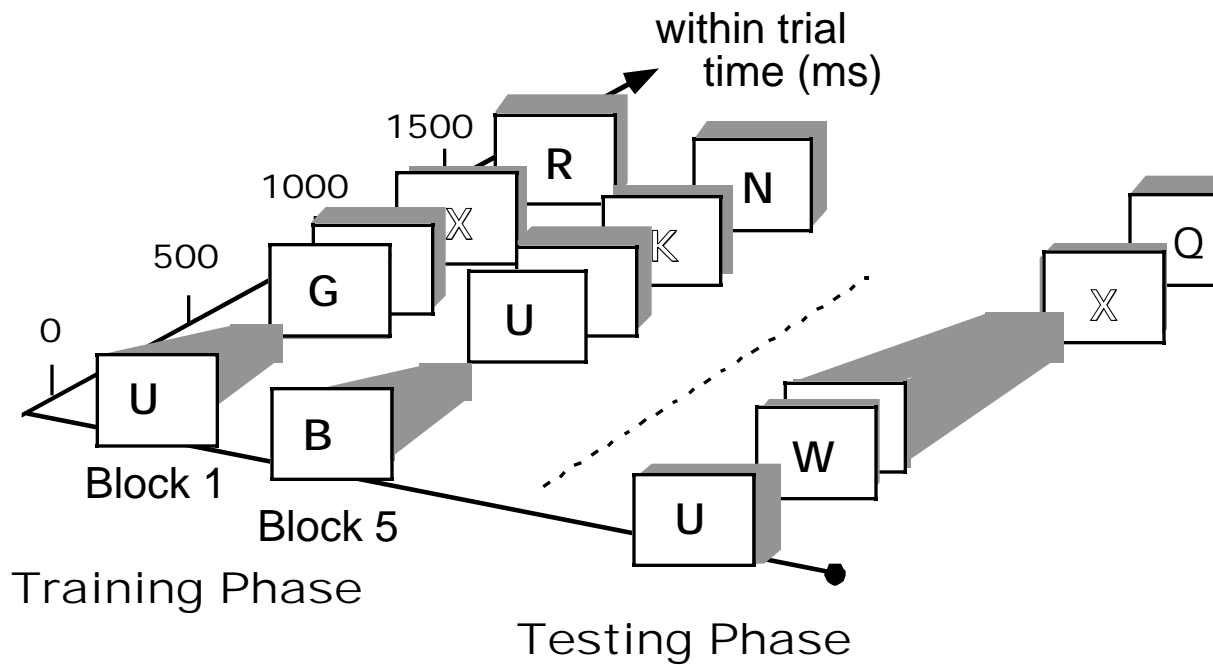


Figure 2

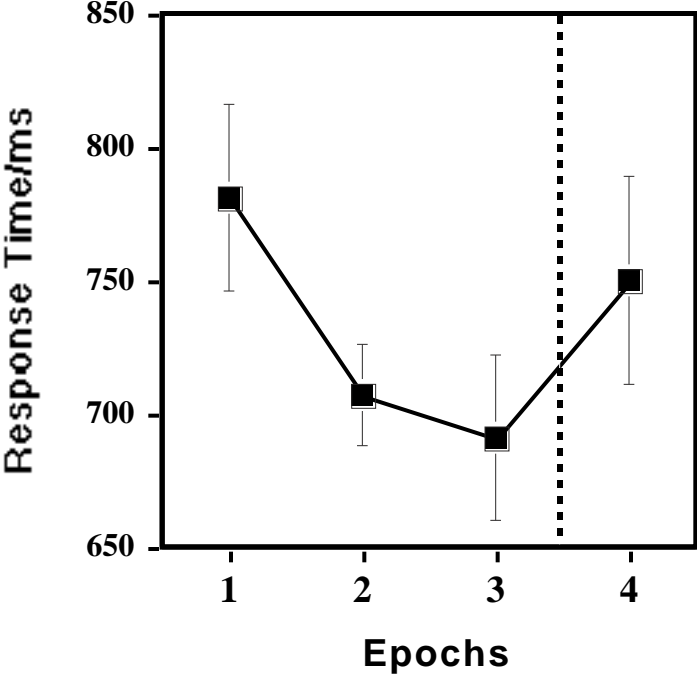


Figure 3

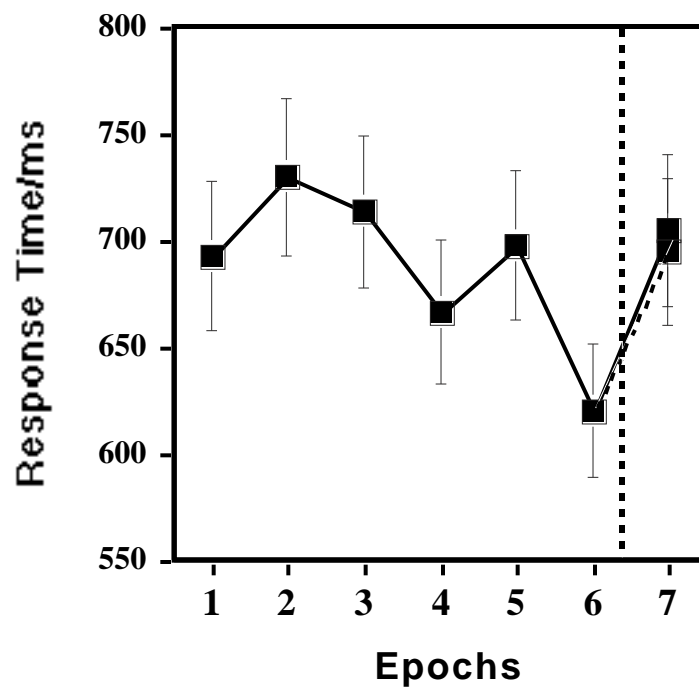


Figure 4

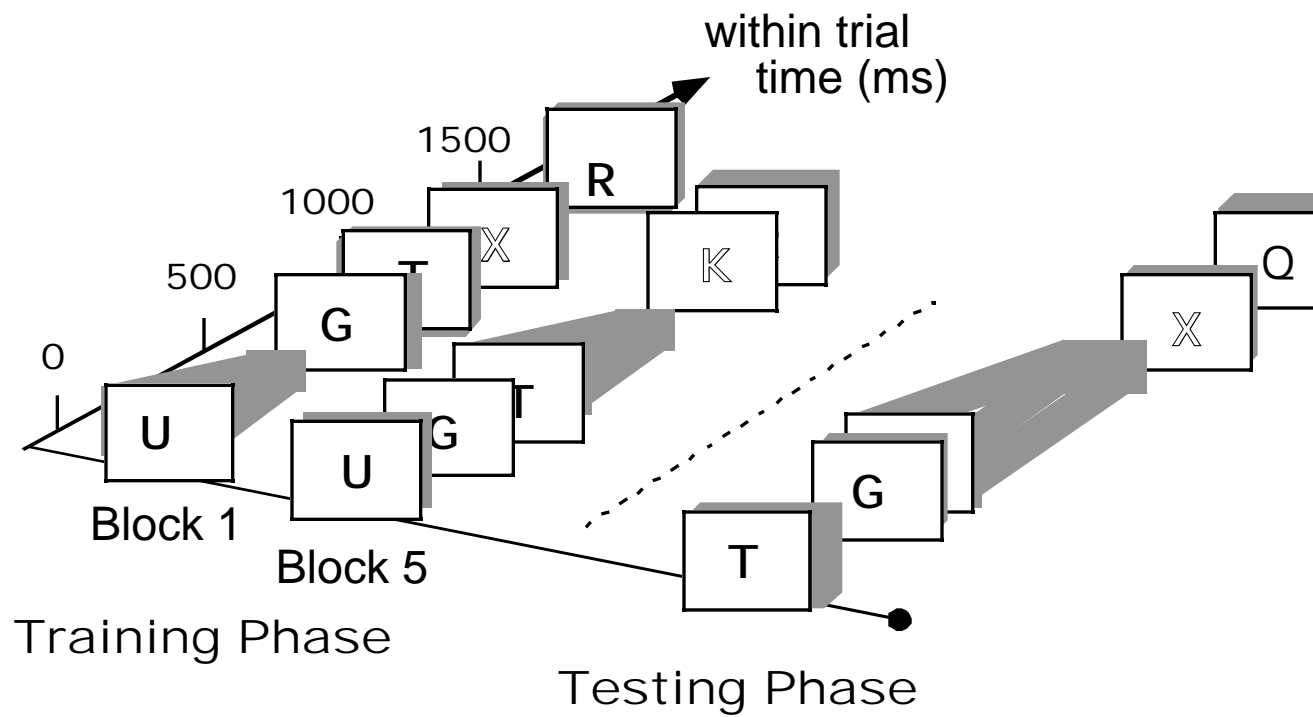


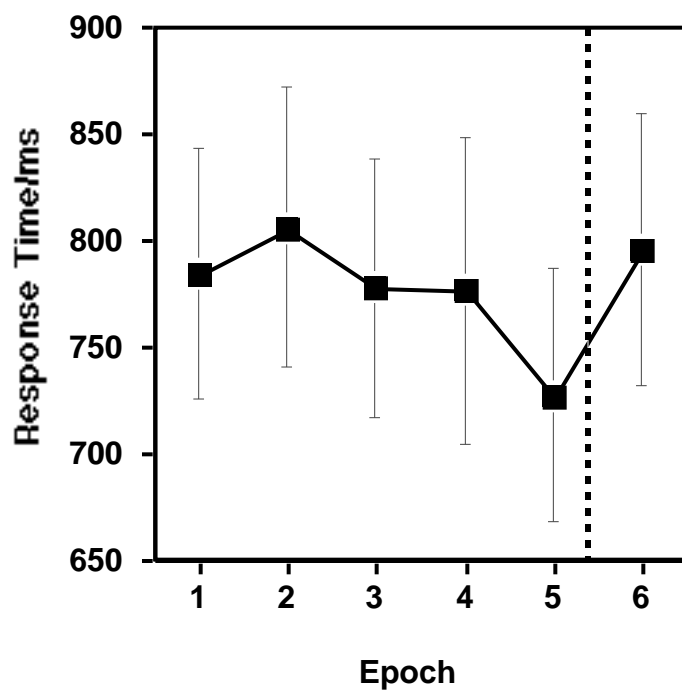
Figure 5

Figure 6

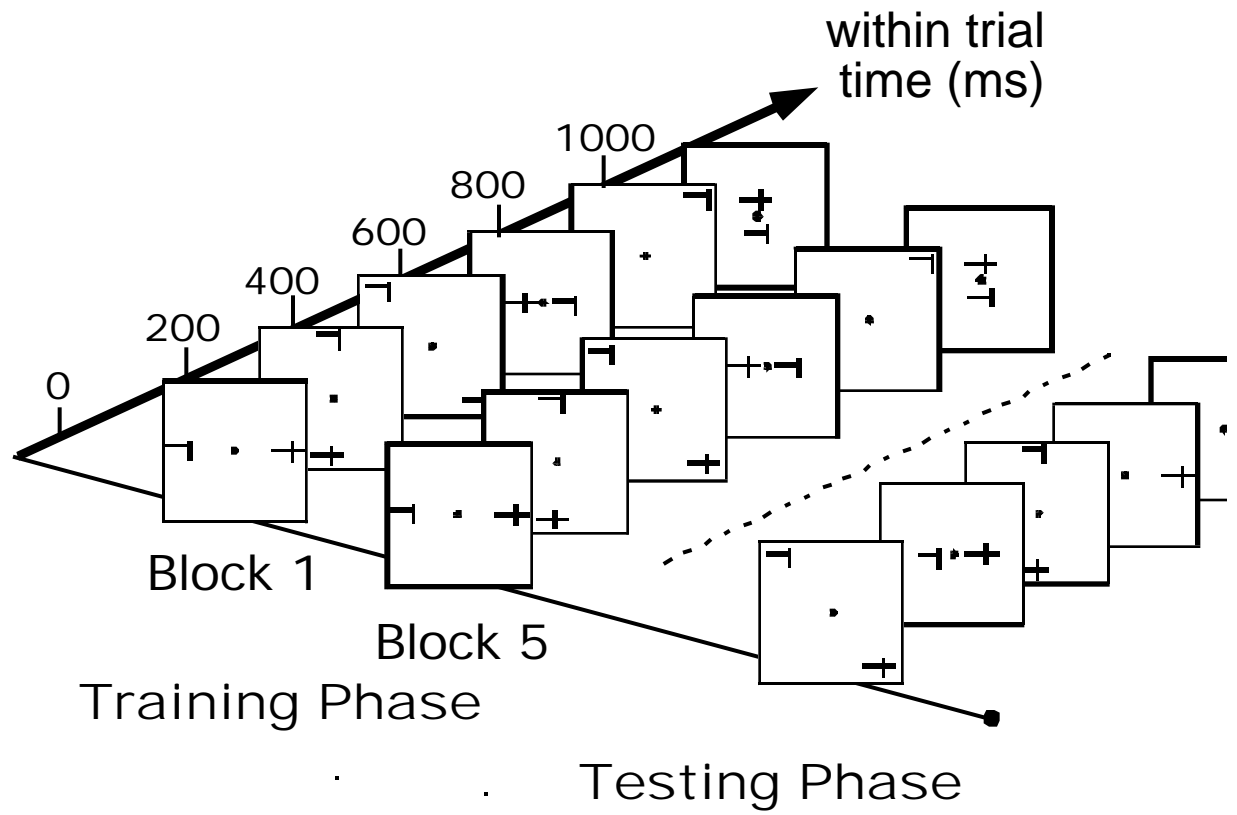


Figure 7

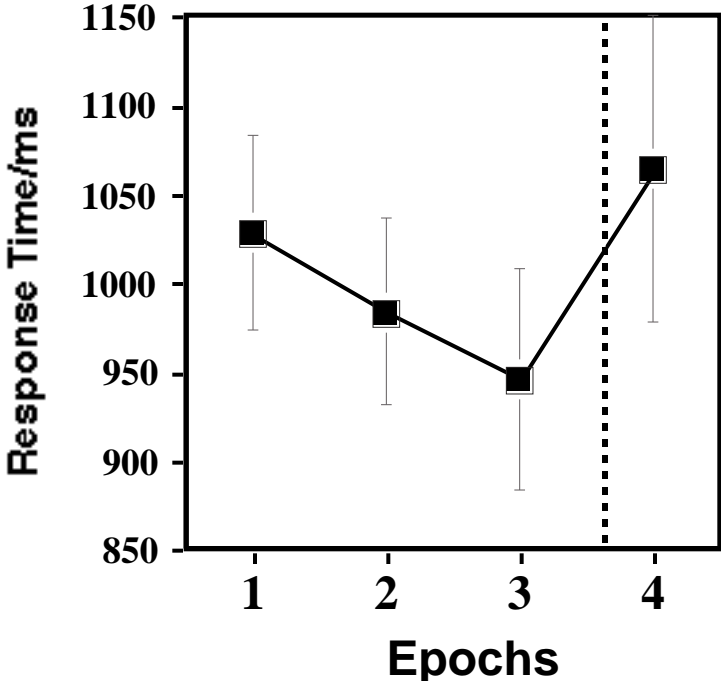


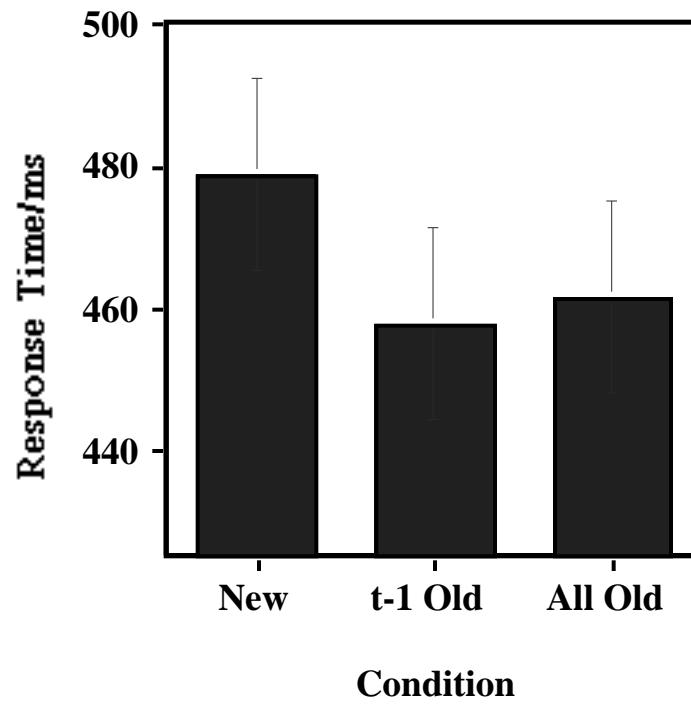
Figure 8

Figure 9