CrossMark

Memory and learning for visual signals in time and space

Sujala Maharjan[1] · Jason M. Gold[2] · Robert Sekuler[1]

Abstract Vision is often characterized as a spatial sense, but what does that characterization imply about the relative ease of processing visual information distributed over time rather than over space? Three experiments addressed this question, using stimuli comprising random luminances. For some stimuli, individual items were presented sequentially, at 8 Hz; for other stimuli, individual items were presented simultaneously, as horizontal spatial arrays. For temporal sequences, subjects judged whether each of the last four luminances matched the corresponding luminance in the first four; for spatial arrays, they judged whether each of the right-hand four luminances matched the corresponding left-hand luminance. Overall, performance was far better with spatial presentations, even when the entire spatial array was presented for just tens of milliseconds. Experiment 2 demonstrated that there was no gain in performance from combining spatial and temporal information within a single stimulus. In a final experiment, particular spatial arrays or temporal sequences were made to recur intermittently, interspersed among, non-recurring stimuli. Performance improved steadily as particular stimulus exemplars recurred, with spatial and temporal stimuli being learned at equivalent rates. Logistic regression identified several shortcut strategies that subjects may have exploited while performing our task.

✉ Robert Sekuler
  vision@brandeis.edu

[1] Brandeis University, Waltham, MA 02453, USA

[2] Indiana University, Bloomington, IN 47405, USA

Introduction

Many cognitive functions build on an ability to register that a previously encountered stimulus has recurred. By manipulating the statistical characteristics within stimuli, researchers have uncovered some key principles of sensory processing and short-term memory. In those studies, detection of recurrence has been explored with auditory, temporal sequences (Julesz 1962; Julesz & Guttman 1963; Guttman & Julesz 1963; Pollack 1971, 1972, 1990), as well as with visual stimuli presented as spatial arrays (Pollack, 1973). More recently, Agus et al. (2010) presented listeners with 1 s long sequences of noise sampled at 44 kHz, and asked them to judge whether the samples in the stimulus' last half replicated the samples in the first half. Overall, listeners performed this challenging task quite well, although success rates did vary among listeners. Subsequently, Gold et al. (2013) adapted Agus et al.'s task to the visual domain. Their subjects saw sequences of eight quasi-random luminances presented at 8 Hz to one region of a computer display. Subjects were instructed to judge whether the final four luminances matched the first four, that is, whether there was a pairwise match between luminances n_1 and n_5, n_2 and n_6, n_3 and n_7, and n_4 and n_8. Performance roughly paralleled what Agus et al. had found with auditory noise. Parallels included not only large individual differences but also evidence of learning with particular sequences that had been preserved ("frozen") and then presented intermittently at random times throughout the experiment. The improved

performance with frozen sequences implicated the formation of some trans-trial memory, whose development was incidental to the subjects' explicit task, which was just to detect within-sequence recurrence (McGeogh & Irion, 1952).

Results from a novel change detection task, (Noyce et al., 2016) delineated vision's and audition's distinct specializations, with vision excelling in spatial resolution, and audition excelling in temporal resolution. Additionally, another recent study showed that task demands dynamically recruit different modality-related frontal lobe regions: a visual task with rapid stimulus presentation activates cortical regions normally implicated in auditory attention, while an auditory task that demands spatial judgements activates regions normally implicated in visual attention (Michalka et al., 2015). These results encouraged us to contrast processing and learning for visual stimuli presented as a temporal sequence, as a spatial display, or, in one experiment, a combination of the two. Although many different tasks and stimuli could have served our purpose, we decided to adapt Gold et al.'s stimuli and task for this purpose. This choice allowed us to build on what they had found, and also to address a question that their paper left unanswered. In the interest of equitable comparisons between spatial and temporal modes of presentation, our subjects made the same kind of judgment with all modes of stimulus presentation. Moreover, stimuli for all modes of presentation were constructed by drawing samples from the same pool of items (random luminances). Even when tasks and stimuli are similar, we expect performance with spatial presentations of visual stimuli to be substantially better than performance with temporal presentations of the same items.

Experiment 1

Using visual stimuli and a task like those in Gold et al., we evaluated the ease with which repetition of luminance subsets could be detected when delivered all at once (as a spatial array) or over time (as a temporal sequence). Of particular interest was the way that performance with spatial stimuli would vary with duration. We reasoned that brief presentations would undermine subjects' ability to carry out the item-by-item comparisons implied by the task instructions, perhaps forcing subjects to fall back to some form of summary statistical representation (Ariely, 2001; Alvarez & Oliva, 2008; Haberman et al., 2009; Albrecht & Scholl, 2010; Piazza et al., 2013; Dubé & Sekuler, 2015). Additionally, we wanted to re-examine Gold et al.'s report that subjects could not detect mirror-image replication of items within temporal sequences of random luminances. Vision's well-documented sensitivity to mirror symmetry made Gold et al.'s result surprising. Unlike the many different kinds of spatial displays in which mirror symmetry is readily detected, mirror symmetry in temporal sequences seemed to be virtually undetectable. However, the cause of that finding is uncertain. It could have resulted from the sequential presentation of stimulus components, from the use of random luminances as stimulus components, or from some combination of the two. So, this experiment included a condition designed to clarify the point. We hypothesized that differences between responses to mirror-image symmetry embedded in spatial displays and responses to mirror symmetry in temporal sequences reflects a difference between how spatial information and temporal information are processed, rather than some singularity of random luminances.

Methods

Subjects

Fourteen subjects, seven female, who ranged from 18 to 22 years of age, took part. In this and our other experiments, all subjects had normal or corrected-to-normal vision (measured with Snellen targets) and each was compensated ten dollars (U.S.). Subjects gave written consent to a protocol approved by Brandeis University's Committee for the Protection of Human Subjects.

Apparatus

Stimuli were generated in Matlab (version 7.10) using the Psychophysics Toolbox extensions (Brainard, 1997). An Apple iMac computer presented the stimuli on a cathode ray tube display (Dell M770) set at 1024×768 pixels screen resolution, and 75-Hz frame rate. A gray background on the display was fixed at 22 cd/m^2. A chin rest enforced a viewing distance of 57 cm. The room was darkened during the experiment. Unless otherwise specified, the conditions just described were maintained across all experiments.

Stimuli

Stimulus luminances were generated using an algorithm from Gold et al. For each temporal stimulus, eight luminances were presented in succession to the same small region of the display, with no break between items. Presentation of a sequence took 1 s. For spatial stimuli, luminances were presented simultaneously, cheek to jowl as a horizontal array. For both types of stimuli, subjects performed the same task, namely, judging whether a subset of contiguous luminances was or was not replicated within the stimulus. For temporal stimuli, subjects were told to judge whether the last four luminances matched the first four; for spatial arrays, subjects were told to judge whether the rightmost four luminances matched the leftmost four.

Luminances were sampled from a Gaussian distribution whose mean, 22 cd/m^2, was equal to the display's steady uniform background. The distribution's standard deviation (8.66 cd/m^2) was supplemented by upper and lower cut-offs that forced possible luminances to fall within the range from 2.33 to 41.67 cd/m^2 (for additional details, see Gold et al.). Note that the small variance among luminances made the items in a stimulus relatively homogeneous. This reduced the likelihood that any one luminance would stand out.

The experimental design entailed five conditions of stimulus presentation: one temporal (Temporal), and four different spatial (Spatial). Each Temporal stimulus comprised eight luminances, presented one after another at 8 Hz to the same square region at the center of the display. Each Spatial stimulus comprised a horizontal array of luminances presented simultaneously around the display's center (see Fig. 1). Although the luminances and timing of Temporal stimuli were identical to those in Gold et al., an item in any stimulus sequence here was 1.25° square, $\sim 4\times$ smaller than in that study. This reduced size ensured that no item in a spatial display would lie no more than 5° to the left or right of fixation.

Each Spatial stimulus was displayed for either 66, 133, or 253 ms. Hereafter, these stimuli are referred to as S_{Shrt}, S_{Med}, and S_{Long} stimuli, respectively. Note that variation in

Fig. 1 Schematic diagram of temporal and spatial stimuli. Note that the Experiment 2 included two additional conditions not shown in the diagram. Both of these can be described as spatio-temporal, as items were presented sequentially, but to adjacent, non-overlapping regions of the display, progressing sequentially, either from left-to-right or from right-to-left

display timing made designations accurate to just ±1 ms. To the three conditions with Spatial stimuli of varying duration, we added a condition in which left and right halves of spatial arrays were mirror reflections of one another. This type of stimulus, which we call Spatial Mirrored (S_{Mirr}), was presented for 66 ms, the same duration that was used for S_{Shrt}. In order to prevent two items of identical luminance from lying adjacent to one another at the center of the array, which would have been a highly distinctive diagnostic feature, S_{Mirr} stimuli comprised just seven square regions instead of eight. Each Spatial array subtended 10° horizontally, while S_{Mirr} arrays were slightly smaller, 8.75°. For all Spatial displays, components were aligned horizontally with no gaps in between.
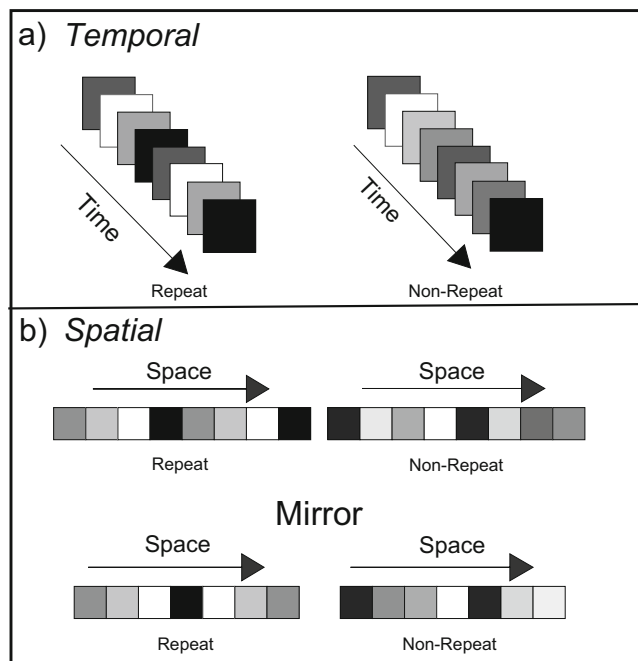
Design

Each subject was tested in all conditions in an order that was counterbalanced across subjects. Subjects completed ten blocks of 110 trials, two blocks for each condition, for a total of 1100 trials per subject. Each block contained equal number of Non-Repeat and Repeat stimuli. The first ten trials in each block were treated as practice, and have been excluded from data analysis. The order of trials within each block was randomized anew for each subject.
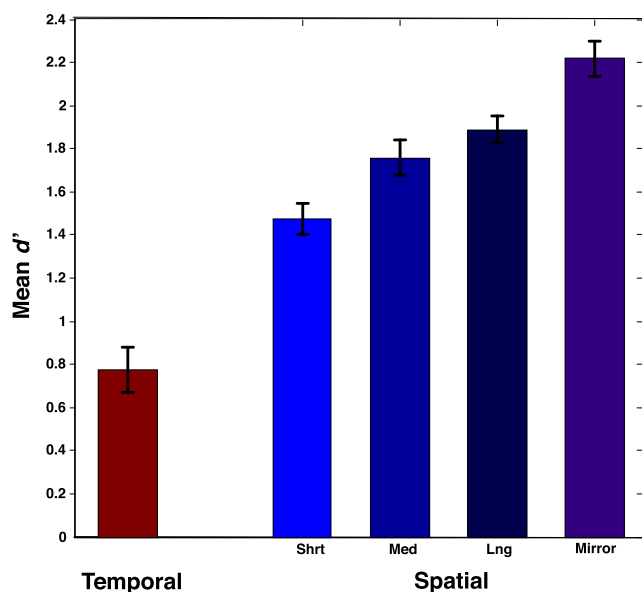
Procedure

After subjects gave written informed consent, they were given a series of diagrams and verbal explanations meant to familiarize them with their task and the types of stimuli they would see. During the experiment, every stimulus was centered on the video display. After each stimulus, a message on the display prompted subjects to press one of two keyboard keys in order to signal whether they thought the stimulus had been Repeat (corresponding luminances matched) or Non-Repeat (corresponding luminances not matched). Immediately after a correct response, a distinctive tone provided feedback.

Results and discussion

We began by evaluating overall performance, expressed as *d'* for the Temporal condition and for the four Spatial conditions, S_{Shrt}, S_{Med}, S_{Long}, and S_{Mirr}. For each block of trials and subject, *d'* was calculated by subtracting *z.pr*(false positives) for Non-Repeat trials from *z.pr*(hits) on Repeat trials. Hits were defined as responses of "Repeated" to Repeat stimuli; false positives were defined as responses of "Repeated" to Non-Repeat stimuli. Figure 2 shows the mean performance in each condition. An analysis of variance (ANOVA) revealed a difference among the five conditions ($F(4,52) = 36.017$, $p < .001$, $\eta^2 = 0.73$). Follow-up tests

Fig. 2 Mean *d'* values from temporal, spatial, and mirror spatial conditions. Error bars are within-subject ±1 SeM

used Bonferroni-adjusted alpha levels of .016 per *F*-test (.05/3) and .025 per *t*-test (.05/2).

Drilling down more deeply into the effect of presentation mode, we compared performance in the three Spatial conditions, S_{Shrt}, S_{Med}, and S_{Long} against performance in the Temporal condition. Detection of repeated items within any of the three types of Spatial stimuli was significantly greater that with Temporal stimuli. This was shown by a repeated-measures ANOVA in which the mean of the three Spatial conditions was contrasted against performance in the Temporal condition ($F(1,13) = 54.337$, $p < .001$, $\eta^2 = 0.81$). A follow up *t*-test showed that even with S_{Shrt}, the briefest Spatial stimulus, performance was better than with Temporal stimuli ($t(13) = 4.76$, $p < .001$, $d = 1.31$). Remarkably, the spatial mode of stimulus presentation produced superior performance even when a spatial stimulus was presented for only $1/60^{th}$ the duration required to present a Temporal sequence. Later, in the General Discussion, we explore possible explanations of this result.

Next, we isolated the effect of duration for Spatial stimuli. An analysis of variance limited to the three Spatial conditions confirmed what can be seen in Fig. 2, namely that performance differed significantly among Spatial conditions of varying duration ($F(2,26) = 6.921$, $p < .01$, $\eta^2 = 0.35$). Follow-on *t*-test showed that the difference between the briefest and the longest durations, that is, S_{Shrt} and S_{Long}, was significant ($t(13) = 4.06$, $p < 0.01$), but the remaining two comparisons were not.

Figure 2 shows that performance was best with S_{Mirr} stimuli. This confirms and extends previous demonstrations that with spatial displays, mirrored (reflectional) symmetry

is detected more readily than are other forms of repetition (Baylis and Driver, 1994; Bruce & Morgan, 1975; Corballis & Roldan, 1974; Barlow & Reeves, 1979; Machilsen et al., 2009; Palmer & Hemenway, 1978; Wagemans, 1997). Our linear arrays, which comprised just a few, relatively large, individual items, differ from displays with which mirror symmetry has been explored previously. As Levi and Saarinen (2004) noted in their study of mirror symmetry detection by amblyopes, "rapid and effortless symmetry perception can be based on the comparison of a small number of low-pass filtered clusters of elements." Once extracted, these distinct clusters would be operated on by some longer-range mechanism that compares clusters located in corresponding positions within a display. The relatively large luminance patches in our displays would have made it easy for vision to isolate the clusters (regions of uniform luminance) that would then enter into a second-stage, comparison process. Because of their proximity, one might imagine that the third and fifth luminance patches (the ones that bookended the middle item) would be most easily compared by the kind of longer-range mechanism that Levi and Saarinen postulated. This led us to ask whether those third and fifth luminance patches made some special contribution, say via selective attention, to subjects' superior performance with S_{Mirr} stimuli.

To test the possibility that with S_{Mirr} stimuli these two luminance patches were especially influential, we simulated performance under two different assumptions with logistic regression. The first model assumed that subjects based their responses to S_{Mirr} stimuli solely on the difference between luminance patches n_3 & n_5; the second model assumed that subjects based their responses on comparisons between items in each corresponding pair of luminance patches, that is, n_1 & n_7, n_2 & n_6, and n_3 & n_5. Note that for all Repeat stimulus exemplars, these two models would always predict exactly the same, error-free performance. That is, no matter what the model, every stimulus would be correctly categorized as Repeat. Therefore, we decided to confine our analysis to Non-Repeat stimulus exemplars, for which the models' predictions would diverge and therefore be more informative. Each logistic regression predicted *pr*(false positive responses) as a function of the difference between the luminances singled out by the model: either only the difference between luminances n_3 & n_5, in one regression, or every difference between corresponding luminance pairs, in the second model. We followed up the regressions with a X^2 difference test on the two nested models. The result showed that the model in which responses to S_{Mirr} stimuli are based only on luminances n_3 & n_5 gave a significantly poorer fit ($X^2(2) = 69.905$, $p < .0001$). It seems unlikely, then, that superior performance with S_{Mirr} stimuli came from selective attention just to luminances n_3 & n_5. Rather, superior performance with S_{Mirr} stimuli might be more

easily understood within the framework that has been proposed for perception of mirror symmetry in other kinds of displays. For example, pre-attentive processes, which have been implicated in mirror symmetry detection (e.g., Wagemans 1997), could be especially potent for displays, like ours, made up of just a few, large elements arranged around a vertical axis at the visual field's center.

Any explanation of the ease with which our subjects detected mirror symmetry begs the question of why Gold et al. found mirror symmetry to be virtually undetectable. In their study, mirror symmetrical stimuli were generated by the same algorithm that we used, but those stimuli were presented sequentially (at 8 Hz), rather than as a spatial array. Before concluding that this difference in results arose from the difference between temporal and spatial presentation modes, we had to rule out the contribution of stimulus size. In particular, each item in our $\mathbf{S}_{Mirr}$ spatial stimuli was one-quarter the size of an item in Gold et al.'s stimuli. Recall that we shrank the stimuli for our experiments so that items in Spatial arrays would not fall too far out in the visual periphery.

To determine whether stimulus size mattered for our results, eight new subjects were each tested on four conditions, temporal mirror and Temporal conditions with stimulus items the same size as those in Gold et al.'s study, and with the same two conditions with stimulus items the same reduced size we used in Experiment 1. Prior to analysis, two subjects' data were discarded because of cell phone use during testing (one confirmed; the other strongly suspected based on his reaction times). Table 1 shows the mean d' values and within-subject standard errors for each condition. For each stimulus size, performance in the Temporal condition was significantly better than the temporal mirror condition. For stimuli with smaller luminance patches ($t(5)$ = 4.04, $p <.05$, $d = 2.55$), and for stimuli with larger patches ($t(5) = 3.17$, $p <.05$, $d = 2.14$). Thus, the poor performance Gold et al. found with mirror symmetrical *temporal* sequences reflected the mode of presentation, not the nature of the items comprising the sequences.

We next asked whether the $\mathbf{S}_{Mirr}$ condition's superior performance came from the fact that, unlike other Spatial stimuli, each $\mathbf{S}_{Mirr}$ stimulus contained just seven luminance patches rather than eight. Perhaps having fewer luminances in a stimulus facilitated detection of a repetition within the stimulus. In order to assess this possibility, we tested three new subjects on $\mathbf{S}_{Mirr}$ and $\mathbf{S}_{Shrt}$ conditions. In both conditions, every stimulus comprised eight luminance patches. For $\mathbf{S}_{Mirr}$ stimuli, the two middle luminances, namely items 4 and 5, were replicates of one another. Table 2 gives the results from these control measures, along with the means and standard errors for the analogous conditions from Experiment 1.

For each subject, d' was appreciably higher with $\mathbf{S}_{Mirr}$ than with $\mathbf{S}_{Shrt}$ stimuli. Moreover, for each subject the difference between stimuli comprising seven items and stimuli comprising eight items was close to the difference found in Experiment 1. This result indicates that the difference between the two conditions in Experiment 1 did not result from the difference in the number of luminance patches in the stimuli. Note that the overall d' values for the three control subjects was less than the corresponding mean value from Experiment 1's subjects. The relatively small standard errors associated with Experiment 1's results leave us at a loss to account for this discrepancy.

Experiment 1 showed a large difference between performance when luminances were presented spatially and when the same luminances were presented temporally. Although the spatial and temporal dimensions of early vision are to some degree separable (Wilson, 1980; Falzett & Lappin, 1983), many psychophysical and physiological results suggests a link between the processing of spatial information and the processing of temporal information (e.g., Doherty, Rao, Mesulam, & Nobre, 2005; Rohenkohl, Gould, Pessoa, & Nobre 2014). These links include a suggestion that information from the two dimensions of processing converges at some site in the parietal lobe (e.g., Walsh 2003; Oliveri, Koch, & Caltagirone 2009). Such convergence might support a combination or coordination of temporal and spatial streams of information, as some psychophysical studies have shown (Goldberg et al., 2015; Keller & Sekuler, 2015). Although the tasks in those studies differed from

Table 1 Mean d' and standard errors from the supplementary experiment

Size of items	Mode of presentation	
	Temporal	Mirror
Smaller	1.22 (0.08)	0.67 (0.06)
Larger	1.21 (0.13)	0.60 (0.06)

Table 2 d' values: Control results and Experiment 1 results

Subject	Mode of presentation	
	$\mathbf{S}_{Shrt}$	$\mathbf{S}_{Mirr}$
1	0.604	1.180
2	0.900	1.601
3	0.543	1.933
Mean	0.690	1.530
Expt. One	1.477	2.218
SeM	0.134	0.144

ours, their results do suggest the possibility that by making both sources of information available at the same time, concurrent spatial and temporal presentations could enhance performance over what would be produced by either source alone. Experiment 2 addressed this possibility, comparing the detection of repetition embedded in spatial, temporal, and spatio-temporal stimuli, in which the two dimensions were combined.

Experiment 2

Experiment 1 showed that detection of repetition within Spatial stimuli was considerably better than it was within Temporal stimuli. In the natural world, many events are characterized not by spatial or temporal information alone, but by some combination of the two. For example, stimulus variation over both space and time contributes importantly to event recognition and understanding (Cristini et al., 2007; Shipley & Zacks, 2008). Experiment 2 examined whether the availability of spatial information and concurrent temporal information would facilitate detection of repetition within a stimulus. One detail of the experiment's design was motivated by the directional bias seen previously, when subjects processed spatio-temporal stimuli (Sekuler, 1976; Sekuler et al., 1973; Corballis, 1996). Specifically, previous studies have shown either a left-to-right or a right-to-left order advantage in processing visual stimuli presented in rapid sequence. As there was no consensus on the direction of the bias, we generated spatio-temporal stimuli with both left-to-right and right-to-left orders of item presentation. So, in addition to assessing the impact of combining temporal and spatial information, we examined how direction of presentation influenced performance.

Methods

Subjects

Fifteen new subjects, seven female, who ranged from 19 to 30 years of age took part. One subject's data were discarded because of extremely low accuracy relative to other subjects.

Stimuli

Stimuli were generated as in Experiment 1. Two conditions, Temporal and S_{Shrt} were brought over identically from Experiment 1. To these, two new conditions were added, which we call spatio-temporal, left-right (S_{lr}) and spatio-temporal, right-left (S_{rl}). In these new conditions, eight square regions varying in luminance were presented successively, with the same timing as in the Temporal

condition. However, unlike the Temporal condition, successive luminances were not delivered to the same region of the display. Instead, for S_{lr} stimuli, luminances were presented successively, the first to a region centered at 5° to the left of fixation, then shifting leftward in steps of 1.25°, and ending at 5° to the right of fixation; in contrast, for S_{rl} stimuli, successive luminances were presented first to a region centered at 5° to the right of fixation, shifting rightward in steps of 1.25°, and ending at 5° to the left of fixation. For both directions of spatio-temporal presentation, the stepwise presentations of luminances spanned 10°, the same horizontal extent as that of a S_{Shrt} stimulus. Each item in S_{lr} and S_{rl} was presented for ~133 ms, the same duration as items in the Temporal stimuli.

Design and procedure

Each subject completed four blocks of 160 trials, one block per condition, in counterbalanced order. Within each block, equal number of Repeat and Non-Repeat stimuli were randomly interleaved (see Fig. 1). As before, the first ten trials of each block were treated as practice, and were excluded from the analysis. Procedures were otherwise identical to those in Experiment 1.

Results and discussion

Figure 3 shows that *d'* varied reliably across stimulus type. This was confirmed by a repeated measures ANOVA ($F(3,39) = 4.073$, $p < .05$, $\eta^2 = 0.24$). More specifically, as Experiment 1 had shown, detection of within-stimulus repetition was better with Spatial stimuli than with Temporal stimuli, even with the briefest Spatial stimuli ($t(13)$
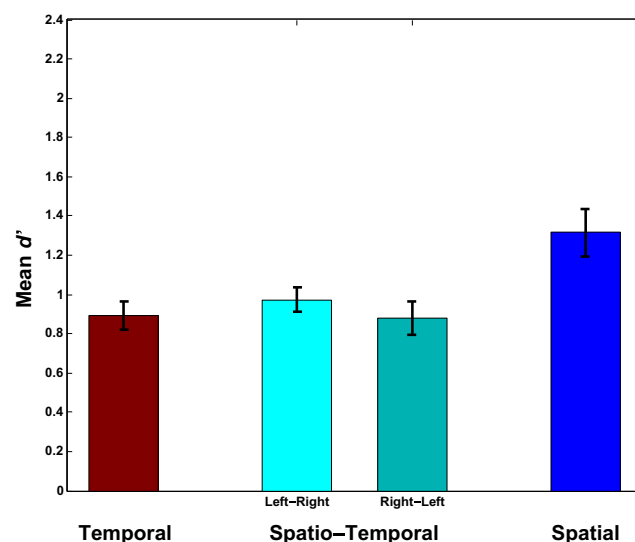

Fig. 3 Mean *d'* values for temporal, spatio-temporal and spatial conditions. Error bars are ±1 SeM

= 2.44, p <0.05, d = 0.61). When spatial information and temporal information were packaged together within a single stimulus (that is, in S_{lr} or S_{rl} stimuli), that combination not only produced performance well below what was seen with spatial information alone, but it even failed to boost performance above that with temporal information. So, there was no evidence that combining the two sources of information, spatial and temporal, aided performance in our task. In fact, that combination actually undermined performance relative to that with spatial information alone. Such poor performance might have come from the challenge of trying simultaneously to perform the novel task, coordinating separate temporal and spatial streams, while also trying to extract some details from each stream (Dutta & Nairne, 1993).

Because stimulus conditions were tested in separate blocks of trials, when tested with either direction of spatio-temporal sequence, subjects could have anticipated the position at which successive items would appear. So, despite instructions to maintain gaze at the fixation point, the predictability of spatio-temporal sequences might have encouraged anticipatory saccades to the position that would be occupied by the upcoming item. Had anticipatory saccades been timed perfectly, spatially distributed successive items would all have fallen on a single region of the retina, converting a spatio-temporal stimulus into one whose retinal images mimicked those of a Temporal stimulus. Even without overt changes in fixation, shifts in spatial attention might have transformed from a spatio-temporal representation to one that was more purely spatial (e.g., Akyürek & van Asselt 2015). However, the substantial difference in results from Spatial and either spatio-temporal condition provides no evidence that such a conversion took place or, if it did, was actually effective.

Additionally, the direction in which spatial information was delivered—left to right or vice versa—was inconsequential ($F(2,26) = 0.464$, $p = 0.63$, $\eta^2 = 0.03$). This null result from direction of presentation, contrasts with previous demonstrations of a directional bias (Effron, 1963; Umiltà et al., 1973; Sekuler et al., 1973; Sekuler, 1976; Corballis, 1996; Effron, 1963; Mills & Rollman, 1980). The failure to find a directional bias with spatio-temporal stimuli might have reflected differences between the task in our experiment and ones used previously. Specifically, our subjects had to judge whether some portion of the sequence repeated or did not repeat; previous studies, which showed a directional bias, required subjects to judge some temporal characteristic of the stimuli, such as order or simultaneity. Additionally, our stimuli were not restricted to one visual hemifield, but crossed the midline. This might have required that attention be distributed between hemifields, thereby dampening differences that might have otherwise
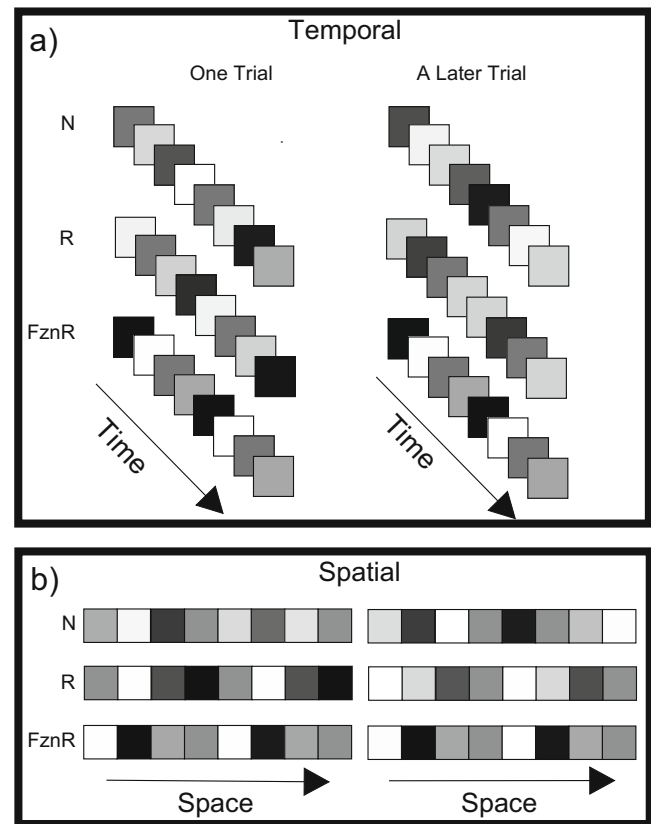
Fig. 4 Schematic diagrams of the three kinds of stimuli presented in Experiment 3. Diagrams in Panel **a** represent stimuli displayed as temporal sequences; diagrams in Panel **b** represent stimuli that are displayed as spatial arrays. Note that an entire sequence of a frozenRepeat stimulus repeats identically in some later trial

been revealed (Awh & Pashler, 2000; Alvarez & Cavanagh, 2005; Delvenne, 2005; Chakravarthi & Cavanagh, 2009; Reardon et al., 2009; Delvenne et al., 2011).

Experiments 1 and 2 showed that repetition of luminances within spatial arrays was more readily detected than was repetition in equivalent temporal presentation. Moreover, this superiority of spatial processing was preserved down even to the briefest spatial presentations we tested. Both experiments focused on within-trial performance, that is, detection of repetition within a stimulus on individual trials. Previous research showed that detection of repetition within temporal presentations of either visual or auditory stimuli could build over trials, giving evidence of learning in both sensory modalities (Agus et al., 2010; Gold et al., 2013). Our third experiment extended those results by repeating the assessment of visual Temporal sequences, and asking whether learning with such sequences was matched by learning with visual stimuli presented as a spatial array.

Experiment 3

With stimuli similar to our Temporal sequences, Gold et al. (2013) found that when the same stimulus exemplar recurred multiple times within a block of trials, subjects' ability to detect a within-sequence repetition of luminance subsets improved for that exemplar. Experiment 3 was intended to replicate that demonstration of learning, and to determine whether the gradual, trial-by-trial improvement with Temporal sequences would also hold for Spatial arrays. We were equally interested to see if trial-over-trial improvement might differ between spatial and temporal stimulus presentations, particularly, whether Spatial stimuli might support more robust, rapid learning. In order to test this, for each subject and condition (Temporal or Spatial), a unique randomly generated Repeat stimulus was stored ("frozen"), and then presented intermittently among the other stimuli. Following Gold et al. and other researchers, we refer to such a stimulus as "frozen" because it is produced by storing and repeating the same Repeat stimulus multiple times throughout a block. Hereafter, we refer to these as Frozen Repeat stimuli (frozenRepeat). On trials with a frozenRepeat stimulus, precisely the same eight items were presented. Such stimuli can be contrasted with Repeat stimuli, which were generated afresh for each presentation. Hereafter, to make this distinction clear, we refer to these stimulus as Fresh Repeat stimuli (freshRepeat).

In order to make equitable comparisons between Spatial and Temporal stimuli, we had to start with stimuli that yielded essentially the same level of performance. Experiments 1 and 2 showed that when presented for as little as 66 ms, a Spatial stimulus produced significantly better performance than a Temporal stimulus (Figs. 2 and 3). To find a duration at which performance with a Spatial stimulus closely matched that with a Temporal stimulus, we carried out a pilot experiment. We tested three experimentally naive subjects with the Temporal stimulus and with Spatial stimuli presented for durations of either 26 or 39 ms, durations ~50% or less than the 66 ms Spatial stimulus. As in the first two experiments, subjects judged whether the luminances in one half of a stimulus did or did not replicate the luminances in the other half. Of the Spatial conditions, stimuli presented for 39 ms produced the d' closest to that of the Temporal condition (0.96 and 0.98, respectively). So, for Experiment 3, Spatial stimuli were all presented for ~39 ms, a duration we expected would produce performance comparable to that with Temporal stimuli.

As just explained, each Spatial stimulus in Experiment 3 was presented for 39 ms, while Temporal stimuli were each presented as in Experiments 1 and 2. For each mode of presentation, Spatial or Temporal, the experimental design included some trials on which a particular Repeat stimulus was made to recur. The purpose was to see whether subjects'

performance would improve with successive encounters with the same Repeat stimulus, not only with Temporal stimuli, as previous research (e.g., Gold et al. 2013; Agus et al. 2010) found, but also with Spatial stimuli, and also to compare the rates of learning in both.

Method

Subjects

Fourteen new subjects participated. They ranged from 19 to 30 years of age, and 12 were female. One subject's data were discarded after it was discovered that she consistently made the same response regardless of the stimulus, that is, her responses were clearly not under stimulus control.

Stimuli

Excepting the shortened duration of Spatial stimuli, stimuli were generated and displayed as in the previous experiments. As before, Temporal and Spatial stimuli were presented in separate blocks of trials. For both Temporal and Spatial stimuli, within each block of trials, precisely the same Repeat stimulus intermittently recurred multiple times (see Fig. 4). Particular fixed repeated (frozenRepeat) stimuli were interspersed among other Repeat stimuli, which were generated afresh for each presentation (freshRepeat stimuli), and also among Non-Repeat stimuli, which were also generated anew for each trial. The particular frozenRepeat stimulus exemplar that recurred was generated anew for each subject and block of trials.

Design and procedure

Each subject was tested on two blocks of Temporal and Spatial stimuli. The order of presentation was counterbalanced across subjects. Each block comprised 220 trials (55 frozenRepeat, 55 freshRepeat, 110 Non-Repeat) trials. As there were three types of stimuli in the current experiment, subjects were allowed more practice than in the previous experiments: here, the first 20 trials were deemed practice. At the start of each block of trials, a unique frozenRepeat stimulus was generated randomly for the subject who would be tested. As a result, if frozenRepeat stimuli did support improvement over trials and that improvement were stimulus selective, as is usually the case for perceptual learning (Sagi, 2011; Hussain et al., 2012; Watanabe and Sasaki, 2014; Harris & Sagi, 2015), using a new frozenRepeat for each block of trials should produce negligible carry over in improvement between blocks. The order of trials was randomized within each block, with the constraint that no two frozenRepeat trials were allowed to occur in immediate succession. This constraint separated successive occurrences of

the same frozenRepeat stimulus by $\bar{M}$ = 3.9 trials (SD = 3.0). Subjects were not told that some stimuli would recur.

Results and discussion

Unlike the previous experiments, in Experiment 3, each subject was tested in two separate blocks of trials for each condition. To verify that results from replications were comparable and that exposure to the frozenRepeat stimuli in the first replication did not affect performance with different frozenRepeat stimuli in the second replication, we examined responses to the first 10 frozenRepeat stimuli in each block. For Temporal frozenRepeat stimuli, hit rates were $\bar{M}$=0.82 and 0.79 for the two replications; for Spatial frozenRepeat stimuli, the hit rates in both replications were $\bar{M}$=0.76. As neither difference between replications was statistically reliable (each $p > 0.05$), we aggregated frozenRepeat results across both replications with each condition.

Figure 5 shows d' values for freshRepeat and frozenRepeat stimuli for Temporal and Spatial conditions. Values were computed using the same method as in the previous experiments. A 2×2 repeated measures ANOVA tested the effects of mode of presentation (Temporal vs. Spatial), stimulus recurrence (frozenRepeat vs. freshRepeat), and the interactions between these variables.

Only recurrence of frozen stimuli produced a significant main effect: $F(1,12) = 19.605$, $p<.001$, $\eta^2 = 0.62$. More specifically, performance with Temporal frozenRepeat stimuli was better than performance with Temporal freshRepeat,
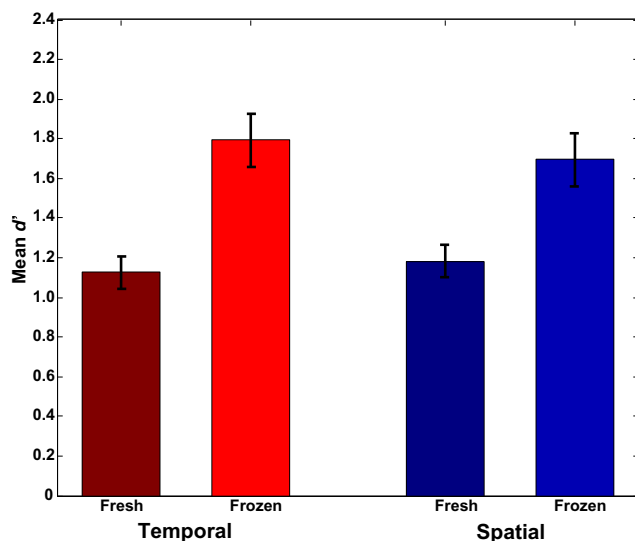

Fig. 5 Mean d' for repeated (freshRepeat) and Fixed repeated (frozenRepeat) stimuli in temporal and spatial modes of presentation. Hit rates for freshRepeat stimuli and frozenRepeat stimuli were calculated relative to false-positive rates for Non-Repeat stimuli. Error bars represent within-subject ±1 SeM. All values are based on averages taken over both blocks of trials for each subject

$(t(12) = 4.53, p <.001, d = 1.17)$. Similarly, performance on Spatial frozenRepeat trials was significantly better than Spatial freshRepeat trials $(t(12) = 3.062, p <.01, d =0.74)$. The two t-tests were conducted using Bonferroni-adjusted alpha levels of .025 per test (.05/2). Mode of presentation, Temporal vs. Spatial, failed to produce a significant effect $(F(1,12) = 0.01, p = .91, \eta^2 = 0.0009)$. The performance advantage enjoyed by frozenRepeat stimuli, for both temporal and spatial modes of presentation, shows that subjects were able to exploit the intermittent recurrence of a particular Repeat exemplar within a block of trials, that is, within each replication. Finally, the interaction between stimulus recurrence and mode of presentation was not significant $(F(1,12) = 0.823, p = 0.38, \eta^2 = 0.06)$. So, despite not forewarning subjects that particular stimulus exemplars might recur from time to time, and despite the fact that subjects' task on any trial did not require recognition that a stimulus had recurred over trials, detection of within-stimulus repetition was superior when the stimulus exemplar was one that recurred on multiple trials.

Figure 5 showed that overall performance with frozen exemplars exceeds performance with fresh ones. We extended this finding in two ways: looking more carefully at individual differences among subjects, and looking at the time course of learning. First, we wanted to know if subjects who benefitted most from multiple encounters of frozenRepeat stimuli in one presentation mode, Spatial or Temporal, also tended to benefit most in the other presentation mode. So, with each subject and with both Spatial and Temporal conditions, we found the mean difference between d' for frozenRepeat and for freshRepeat stimuli. Positive values of this difference indicate an advantage of frozenRepeat over freshRepeat stimuli. As Fig. 6 suggests, most but not all subjects showed an advantage of frozenRepeat over freshRepeat, and did so for both stimulus presentation modes.

To quantify the relationship shown in that figure, we computed the Pearson correlation between results for the two modes. This produced $r = 0.43$, with $p <.07$ (one-sided). Although it falls short of statistical significance, this correlation suggests that a subject who benefits from recurrence of frozenRepeat stimuli in one mode, either Temporal or Spatial, tends also to benefit from recurrence in the other mode. The modest, non-significant correlation between learning with Temporal and Spatial stimuli could implicate a cognitive strategy in which subjects learn some template or pattern, and then use it to recognize when a frozenRepeat stimulus repeats. As the weak, but non-zero correlation suggests, some subjects are better at this than others. Moreover, templates for Temporal and for Spatial stimuli seem likely to take different forms, and to recruit non-identical neural substrates (Noyce et al., 2016; Serences, 2016).
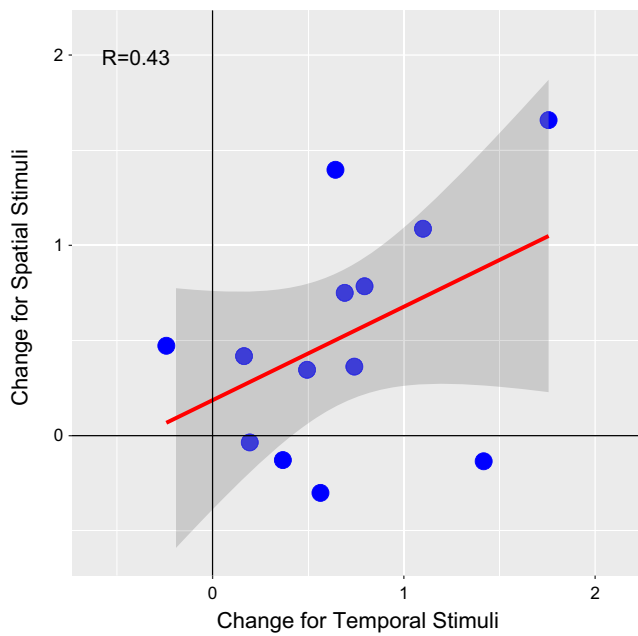
Fig. 6 Mean improvement with frozenRepeat Spatial stimuli versus that for frozenRepeat Temporal stimuli. Data points are for individual subjects; also shown are the maximum likelihood line of best fit, and a ribbon representing 95 % confidence intervals around the best fit line

Figure 5 shows that both modes of presentation, Temporal and Spatial, supported comparable overall levels of improvement with frozenRepeat stimulus exemplars. Figure 7 shows the rate of learning over trials for each mode of presentation. The figure's left and right panels show hit rates for frozenRepeat and freshRepeat stimuli,

respectively. To generate Fig. 7, Temporal frozenRepeat trials and Spatial frozenRepeat trials were separately averaged within successive sets of five trials. This process, carried out for Temporal as well as Spatial stimuli, produced 11 sets of trials per stimulus type. The figure plots the proportion of correct *pr*("Repeat") responses ("hits") against the ordinal number of the mean trial within each successive set of trials.

With frozenRepeat stimuli, repeated intermittent presentation of either Spatial stimuli or Temporal stimuli produced a gradual increase in hit rate. To quantify this evidence of learning, we bootstrapped the data for each of the four stimulus types: the freshRepeat and frozenRepeat varieties of both Spatial and Temporal stimuli. From each bootstrap sample, we computed the slope of the best-fit linear function linking *pr*("Repeat") judgments to the ordinal position that a stimulus occupied within a block of trials. If learning were stimulus selective, that is, if it were most pronounced for recurring stimuli, slopes for frozenRepeat stimuli would be positive and would exceed those for freshRepeat stimuli. To test this prediction, we found the mean and 95 % confidence intervals from 1,000 bootstrapped linear fits for each of the four conditions. For frozenRepeat stimuli, the means of the bootstrapped slopes were reliably greater than zero, $\bar{M} = 0.0017$, 95 % CI [0.0017, 0.0018] and $\bar{M} = 0.0011$, 95 % CI [0.0010, 0.0012] for Spatial and Temporal stimuli, respectively. In contrast, for freshRepeat stimuli, mean slopes were $\bar{M} = 0.00055$, 95 % [.00049, 0.00068] and $\bar{M} = -0.00056$, 95 % CI [-0.00065, -0.00036] for Spatial and Temporal stimuli respectively. Neither of these latter values was reliably different from zero, the value expected had there been no learning. These results confirm that subjects'
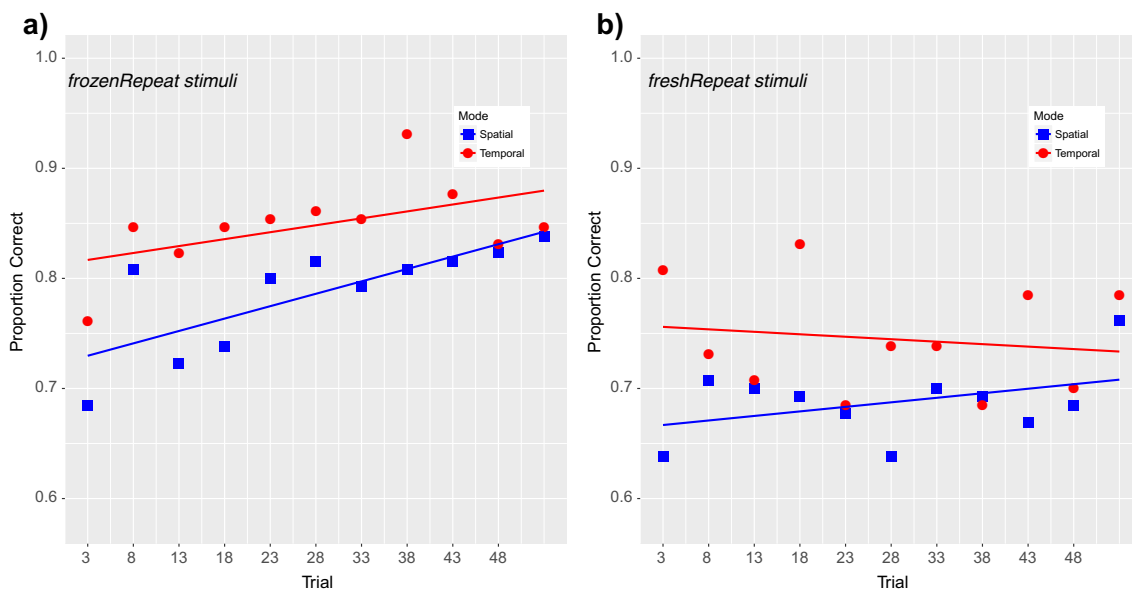
Fig. 7 Trial-wise performance in Experiment 3. Mean proportion correct "Repeat" judgments are plotted as a function of the mean trial within successive sets of five trials. Panel **a** shows results for frozenRepeat stimuli. Panel **b** shows results for freshRepeat stimuli. Within each panel data for Spatial (■) and Temporal (●) stimuli are shown separately

performance with frozenRepeat stimuli shows a reliable improvement with successive trials. Moreover, trial-over-trial performance with frozenRepeat stimuli increased in nearly equal measure with both spatial arrays and temporal sequences.

General discussion

Each of our three experiments measured ability to detect repetition of a subset of elements within a stimulus made up of random luminances. Stimuli were presented either as temporal sequences, as spatial arrays, or as sequences presented in both time and space. Subjects were generally better at detecting repetition within a stimulus whose components were presented as a spatial array rather than as a temporal sequence. In fact, if one wanted to equate performance between the two modes, one would have to limit the display of spatial arrays to just ~39 ms. Additionally, combining spatial and temporal features failed to enhance performance over either presentation mode, as one might have expected when information from separate sources was combined. Instead, performance with combined sources was limited to the level seen with Temporal stimuli alone, the presentation mode that produced the poorest performance. Of course, we are mindful that this result could have been altered by prolonged practice with the unusual demands of the spatio-temporal stimuli. Furthermore, when luminances comprising one side of a spatial array were mirror reflected on the other side, detection of repetition was best. This differs from Gold et al. (2013)'s result with sequentially presented mirror symmetrical luminances.

Perceptual learning and task demands

In Experiment 3, random intermittent presentation of the same stimulus exemplar boosted detection of repetition within that stimulus, and did so to about the same degree for both Spatial and Temporal modes of presentation. That perceptual learning would be comparable for the two modes of presentation was not inevitable. After all, though performance on Spatial and Temporal stimuli had been equated and though subjects achieved almost the same level of final performance with frozenRepeat stimuli of both types, the two modes of presentation entail different task demands, not only in Experiment 3, but in all of our experiments.

To appreciate this point, consider what a subject might do in deciding whether items were repeated within a temporal presentation. The task instructions imply that a subject might separately encode and store each of a sequence's first four luminances, $n_1 \ldots n_4$, and then compare the memory of each against its corresponding luminance, that is, compare n_1 to n_5, n_2 to n_6, and so on (Kirchner, 1958). Presumably,

if a sufficiently strong mismatch signal resulted from any one of these comparisons, the appropriate response would be "No Repeat"; if no such mismatch signal resulted, the appropriate response would be "Repeat". Note that various sources of noise, including masking and rapid adaptation (Crawford, 1947; Neisser, 1967; Breitmeyer, 2007), would introduce variability in the perception of any item. As a result, a subject who responded "No Repeat" whenever corresponding items seemed to differ by even the smallest amount, would certainly reject Repeat stimuli much of the time. Therefore, a prudent subject would establish a criterion for deeming stimuli as Repeat stimuli if its two halves differed by some small amount, rather than by zero. The task demands imposed by Temporal stimuli suggests one reason why detecting mirror symmetry in such stimuli might be so difficult, as shown by Gold et al. and confirmed by us. After all, detecting mirror symmetry in a sequential presentation might require that individual items held in working memory be reordered or perhaps mentally rotated before comparisons were made. Such challenging and error-prone operations could explain the difficulty subjects had in detecting mirror symmetry embedded in temporal sequences.

In contrast to the serial processing demanded by all Temporal stimuli, their Spatial counterparts allowed a different approach. In particular, because each spatial array was limited in size (just ±5° of fixation) and because each item within an array was sufficiently large, spatial arrays lent themselves to parallel processing (Sekuler & Abrams, 1968; Cunningham et al., 1982; Hyun et al., 2009).

Taking short-cuts?

As explained earlier, the instructions that subjects received encouraged them to base decisions on detailed, item-by-item comparison of corresponding luminance levels. In order to comply with these instructions, all the individual luminances and their order within a stimulus would have had to be extracted, held in memory, and then operated on. However, subjects certainly would have enjoyed a fair measure of success had they adopted some shortcut, basing their decisions on a truncated or sparse representation of the stimulus. For example, the algorithm used to construct our stimuli guarantees that subjects would have greater than chance success if their judgments were based on, say, just two or three luminances, rather than all eight. Moreover, subjects have been known to fall back onto short cuts when conditions make it impractical to do otherwise, e.g., when stimuli are complex and/or are presented very briefly (Haberman et al., 2009).

Our approach was to predict performance based on different sets of assumptions about the information that subjects used to make their decisions about whether a stimulus was Repeat or Non-Repeat. The complexity of our

stimuli and task made many different shortcuts possible, which put an exhaustive search of shortcuts out of reach. Instead, we opted for a shortcut approach—examining just three different sets of assumptions about the information subjects used.

In doing this, we focused exclusively on trials with Non-Repeat stimuli, determining how well each of three different shortcuts might account for the correct rejections ("Non-Repeat" responses) and the false positives ("Repeat") made to such stimuli. Our focus had to be limited to Non-Repeat trials because modeling Repeat trials would have required some explicit statement of the noise level associated with each item in a stimulus. As mentioned earlier, if one assumed zero noise, any model would wrongly predict perfect, error-free performance on every trial with an Repeat stimulus. Consequently, we limited our assay to Non-Repeat trials. Additionally, we narrowed our focus to Non-Repeat trials from Experiment 1. That experiment had the largest range of different conditions, and we were particularly interested in the possibility that subjects' dependence on shortcuts might vary with condition. For example, would differences in results from various Spatial conditions reflect differences in reliance on shortcuts?

Using binomial logistic regression (*R*'s *glm* routine with a logit link), we evaluated how well each of three sets of assumptions predicted subjects' performance on Non-Repeat trials. Each predictor represented different assumptions about how subjects used a stimulus' luminances to classify a stimulus as Repeat or Non-Repeat. One predictor embodied what subjects would have done had they followed the instructions they had been given, basing their responses on all the information in a stimulus. The other two predictors represented shortcuts in which different amounts of stimulus information was omitted. Note that because we worked only with Non-Repeat trials, all "Non-Repeat" responses qualified as correct rejections, and "Repeat" responses as incorrect, false positives.

The three predictors

One predictor, called *fullInfo*, assumed that subjects complied with the instructions they received at the outset of testing. *fullInfo* assumed that responses were based on the absolute difference between each of the four corresponding luminance pairs, n_1 and n_5, n_2 and n_6, n_3 and n_7, and n_4 and n_8. There are several different operations that could have been applied to the four pairwise differences, once they had been extracted from the stimulus and entered into memory. These alternative operations are closely related to one another, and would all produce very similar results. For our logistic regressions, we assumed that subjects based their responses on the largest of the four pairwise differences. Following Sorkin's (1962) differencing model for

"same"-"different" judgments, the predictor assumed that subjects compared the magnitude of this maximum difference between corresponding items against a criterion value. If that difference exceeded the criterion, the stimulus was categorized as "Non-Repeat", otherwise, as "Repeat".

The remaining two predictors agree in assuming that some information contained in each Non-Repeat stimulus fails to enter into subjects' decisions. However, these two predictors differ on exactly how much information is discarded. One predictor, *sumStats*, assumes that quite a lot of information is discarded. Specifically, *sumStats* assumes that subjects discard nearly all item-order information. Rather, subjects are assumed to sum all the luminances within each half of stimulus. The result is a pair of scalars, one representing the first half of the stimulus, the other, the second half. The decision of "Repeat" or "Non-repeat" then tracks the absolute difference between the resulting pair of scalars, summary representations of each half stimulus. This is not a new idea. In fact, this same predictor successfully accounted for much of subjects' ability to categorize rapidly presented temporal sequences of varying luminance (Gold et al., 2013), as well as multisensory temporal sequences whose varying luminances were accompanied by concurrent tone that varied in frequency (Keller and Sekuler, 2015). We expected that *sumStats* would be effective in predicting responses to our Non-Repeat stimuli that were presented temporally, as in those prior studies.

In terms of what information enters into subjects' decisions, our final predictor, *xTremes*, is intermediate between *fullInfo*, in which all information is used, and *sumStats*, in which minimal information is used. Specifically, *xTremes* assumes that a subject identifies the most deviant item within each half stimulus, and bases a response on just the absolute difference between these two extreme items. With Non-Repeat stimuli, the probability of a "No Repeat" response would increase as the difference between the two extremes increases. We hypothesized that this predictor would be particularly influential with Temporal sequences because the temporal structure of these stimuli could be transformed into an quasi-auditory code, as Guttman et al. (2005) demonstrated for rapidly varying visual stimuli. The temporal structure of our Temporal stimuli might make it easier to pick out the extremes in luminance, much as extremes are readily detected in an auditory contour of varying pitches or loudnesses (Green et al., 1984).

Applications of logistic regression

To evaluate how well the three predictors accounted for subjects' performance, we carried out one multiple logistic regression on trials from each stimulus condition. The model for each logistic regression included all three predictors, *fullInfo*, *sumStats*, and *xTremes*. Interaction terms were

not included in the regression. In each logistic regression, all subjects' data were pooled. Finally, before performing logistic regressions, we compensated for differences in the three predictors' ranges of output values (dependent variables). To assess statistical significance we adopted a Bonferroni-adjusted alpha level of .003 per test (that is, .05/15).

Table 3 summarizes the results for each predictor and stimulus condition. The table also shows the regression coefficients and their associated *p* values. Cells representing statistically significant results are highlighted in blue. Also shown in the table are the false positive rates associated with Non-Repeat stimuli in the different conditions, and their corresponding bootstrapped 95 % confidence limits. For each stimulus condition, the overall regression model (not shown in the table) was well fit: the *p* value from the X^2 test for each fit was *p* <0.0001.

Consider first the results with *fullInfo*, the predictor that was closest to what subjects were instructed to do, that is, evaluate and compare all corresponding luminance pairs. Surprisingly, *fullInfo* turned out to significantly predict subjects' performance in just one condition, S_{Mirr}. Although we cannot say definitively that subjects generally failed to comply with the instructions they received, this result suggests a rather weak connection between those instructions and what subjects actually did. Likely, this weak connection reflects the challenge of actually complying with those instructions,

plus the ease with which correct judgments could be made on other, simpler bases.

Consider next the results when *sumStats* is used as a predictor. Earlier, we cited two studies in which *sumStats* was a good predictor of performance with Non-Repeat stimuli that were presented as temporal sequences, at 8 Hz. Table 3 confirms those previous results: here, too, *sumStats* is a significant predictor of responses to Temporal stimuli. The table shows that *sumStats*' predictive power extends beyond Temporal stimuli, to all four varieties of Spatial stimuli. So, *sumStats* turns out to be an effective shortcut across the board, in every condition.

Finally, *xTremes* proved to be a significant predictor for every Spatial condition except S_{Mirr}. Contrary to what we predicted, *xTremes* failed to significantly influence performance with Temporal stimuli. This unexpected result suggests that even if rapid temporal variation of a visual stimulus were transformed into some auditory-like code, as Guttman et al. suggested, extracting and then using the extremes from such a code is not a sure thing. An actual explanation for the failure of our prediction will require additional work.

Table 3's pattern of results is both interesting and surprising. It is interesting in that *sumStats*, the predictor that incorporates the least amount of stimulus information, does the best job overall of predicting performance. It is surprising in that the most limited success came with *fullInfo*, the

Table 3 Logistic regression results, including slope estimates and *p* values, along with proportions of false-positive responses

| Predictor | Stimuli | | | | |
	S_{Shrt}	S_{Med}	S_{Long}	S_{Mirr}	Temporal
fullInfo					
estimate	0.307	0.029	-0.696	-1.759	-0.653
p	0.509	0.946	0.105	7.12e-05	0.147
sumStats					
estimate	-1.925	-2.016	-1.988	-1.988	-1.843
p	1.45e-06	2.23e-06	2.02e-05	8.12e-06	3.66e-07
xTremes					
estimate	-1.723	-1.683	-1.525	-0.974	-1.076
p	0.0001	3.42e-05	9.61e-06	0.009	0.001
False positives					
proportion	0.29	0.25	0.22	0.22	0.44
95% CIs	[.27, .31]	[.23, .27]	[.20, .24]	[.21, .24]	[.41, .46]

Note: Cells overlaid with blue exceed the Bonferroni-corrected significance level of 0.003. Confidence intervals for proportions of false positives are each based on 1,000 non-parametric bootstrap samples

predictor whose operations most closely paralleled those implied by the instructions, and was based on all the information in a stimulus. As with any regression, the results in the table, no matter how interesting and surprising they may be, cannot pin down exactly what causal influences are at work. Moreover, the fact that some predictor is statistically significant does not mean that subjects rely upon it to the exclusion of other predictors. Our logistic regression models were designed to characterize the relative importance of the three predictors, each assessed separately. Obviously, this formulation was a simplification, but a necessary one, in our view. In particular, a full model would have included not only the three individual predictors, but also six two-way interaction terms and a three-way interaction term. The outcome of such models would have been extremely difficult to interpret in the absence of clear *a priori* ideas about those interactions.

Finally, consider one more aspect of the regression analyses. As explained earlier, our regression-based modeling focused only on trials whose stimuli were of the Non-Repeat variety, necessarily leaving aside an examination of trials on which Repeat were presented. Many of our other analyses represent performance in terms of *d'* values, which are based, of course, on responses on both Repeat and Non-Repeat trials. In contrast, only Non-Repeat trials were included in the regression analyses. However, we do not think that excluding Repeat trials created an absolute barrier to linking regression results to results expressed as *d'* values. In particular, unless a subject knew ahead of time whether a stimulus would be Non-Repeat or Repeat, it would be logically impossible to apply some shortcut on Non-Repeat trials but not on Repeat trials. So, if a regression analysis suggests that some particular shortcut is exploited on Non-Repeat trials, that same shortcut would have had to be used on Repeat trials, too.

Limitations and links to other paradigms

Although we varied stimulus duration for Spatial stimuli, duration and rate of presentation for Temporal stimuli were fixed at the values used previously by (Gold et al., 2013) As a result, we cannot rule out the possibility that altering duration and/or rate would have altered performance with Temporal stimuli. The presentation rate for Temporal stimuli was chosen to be near the peak of vision's temporal modulation transfer function (Kelly, 1961). Additionally, exploratory tests showed that higher rates of presentation reduced performance substantially, presumably by increasing the difficulty of separately processing individual items in a Temporal stimulus. The choice of stimulus presentation rate constrained the duration of a stimulus. In particular, we settled on a one-second duration so that subjects would not

have to store more than four items from each half-stimulus, thereby controlling the burden on working memory (Saults & Cowan, 2007).

Note that had we increased the duration of Spatial stimuli and/or decreased the number of luminances in each Temporal stimulus in Experiment 1, performance with Temporal stimuli could have aligned with performance with Spatial stimuli. Instead, in order to equate performance with the two types of stimuli in Experiment 3, we reduced the duration of Spatial stimuli. Additionally, there are other stimulus variables that seem to be worthwhile exploring in future studies, for example, the variance of distributions from which stimulus luminances were drawn. Higher variance among the luminances of items within a stimulus would have produced more extreme luminance values. These extreme values, in turn, could have served as cues to whether a stimulus was Non-Repeat or Repeat.

Our spatial arrays were sufficiently restricted in size (just $\pm5°$ around fixation) that they could have been processed in parallel, without fixations during a stimulus. Of course, it may be that with our longest Spatial stimuli, subjects could have initiated and completed one saccade while the stimulus was still visible. The latency for most medium amplitude saccades (5–10°) is ∼200 ms and their durations are just a few tens of milliseconds (Carpenter, 1988; Kowler, 2011). Even though subjects could have made one saccade, shifting fixation, during a stimulus presentation in the S_{Long} condition, that additional fixation is unlikely to explain performance relative to the Temporal condition. After all, the two shorter Spatial conditions, whose durations foreclose the possibility of more than one fixation still produced better performance than the Temporal condition.

Finally, some of the processes that seem to underlie successful performance in our experiments may also be implicated in category representation and category learning more generally (Ashby & Maddox, 2005). An abstract description of our stimuli may help to illustrate this point. Each of our eight-item stimuli can be represented as a vector in an n-dimensional space, where n is the product of 8 (the number of ordinary positions in a sequence or in non-mirrored spatial array) and the number of perceptually discriminable levels over our luminance range (2.33 to 41.67 cd/m^2). Although we do not know the precise number of discriminable levels, existing data can be used to estimate it (Graham & Kemp, 1938). Estimating the Weber fraction conservatively, as 0.10, the luminance range for our stimuli comprised ∼30 perceptually distinct levels, ignoring complications from forward and backward masking interactions within a stimulus. As a result, a complete representation of our stimuli could require a space as large as ∼ 8×30 dimensions.

Unlike most stimuli used to study category formation and category learning, the stimuli that populated our two

categories, Repeat and Non-Repeat, are completely inter-mingled in a shared high-dimensional space. As a result, no simple, low-dimensional decision bound could cleanly separate the two categories. This would have made category formation difficult within our stimulus space. Additionally, because of the algorithm used to generate our stimuli, potential Repeat stimuli are far less numerous than Non-Repeat stimuli, and unlike Non-Repeat stimuli, which tile the space densely, Repeat stimuli are distributed as isolated points in that high-dimensional space. As a result, the paradigm defined by our task and stimuli differs from ones, such as the rule-based learning or information-integration tasks that are often used to study categorization (Ashby & Maddox, 2011). Arguably, though, the richness and complexity of our task and stimuli could be useful vehicles in the future for expanding the framework of research on category representation and learning.

References

Agus, T.R., Thorpe, S.J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, *66*(4), 610–618.

Akyürek, E.G., & van Asselt, E.M. (2015). Spatial attention facilitates assembly of the briefest percepts: Electrophysiological evidence from color fusion. *Psychophysiology*, *52*, 1646–1663.

Albrecht, A.R., & Scholl, B.J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567.

Alvarez, G.A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.

Alvarez, G.A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398.

Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Ashby, F.G., & Maddox, W.T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.

Ashby, F.G., & Maddox, W.T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161.

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 834–846.

Barlow, H.B., & Reeves, B.C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, *19*(7), 783–793.

Baylis, G.C., & Driver, J. (1994). Parallel computation of symmetry but no repetition within single visual shapes. *Visual Cognition*, *1*(4), 377–400.

Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Breitmeyer, B.G. (2007). Visual masking: past accomplishments, present status, future developments. *Advances in Cognitive Psychology*, *3*(1-2), 9–20.

Bruce, V.G., & Morgan, M.J. (1975). Violations of symmetry and repetition in visual patterns. *Perception*, *4*, 239–249.

Carpenter, R.H.S. (1988). *Movements of the eyes*, 2nd ed. London: Pion.

Chakravarthi, R., & Cavanagh, P. (2009). Bilateral field advantage in visual crowding. *Vision Research*, *49*(13), 1638–1646.

Corballis, M.C. (1996). Hemispheric interactions in temporal judgments about spatially separated stimuli. *Neuropsychology*, *10*, 42–50.

Corballis, M.C., & Roldan, C.E. (1974). On the perception of symmetrical and repeated patterns. *Perception & Psychophysics*, *16*(1), 136–142.

Crawford, B.H. (1947). Visual adaptation in relation to brief conditioning stimuli. *Proceedings of the Royal Society of London Series B, Biological Sciences*, *134*(875), 283–302.

Cristini, M., Bicego, M., & Murino, V. (2007). Audio-visual event recognition in surveillance video sequences. *IEEE Transaction on Multimedia*, *9*(2), 257–267.

Cunningham, J.P., Cooper, L.A., & Reaves, C.C. (1982). Visual comparison processes: identity and similarity decisions. *Perception & Psychophysics*, *32*(1), 50–60.

Delvenne, J.-F. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition*, *96*(3), 79–88.

Delvenne, J.-F., Castronovo, J., Demeyere, N., & Humphreys, G.W. (2011). Bilateral field advantage in visual enumeration. *PLoS One*, *6*(3), 1–8.

Doherty, J.R., Rao, A., Mesulam, M.M., & Nobre, A.C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *Journal of Neuroscience*, *25*(36), 8259–8266.

Dubé, C., & Sekuler, R. (2015). Obligatory and adaptive averaging in visual short-term memory. *Journal of Vision*, *15*(3), 1–13.

Dutta, A., & Nairne, J.S. (1993). The separability of space and time: dimensional interaction in the memory trace. *Memory & Cognition*, *21*(4), 440–448.

Effron, R. (1963). The effect of handedness on the perception of simultaneity and temporal order. *Brain: A Journal of Neurology*, *186*, 261–284.

Falzett, M., & Lappin, J.S. (1983). Detection of visual forms in space and time. *Vision Research*, *23*, 181–189.

Gold, J.M., Aizenman, A., Bond, S.M., & Sekuler, R. (2013). Memory and incidental learning for visual frozen noise sequences. *Vision Research*, *99*, 19–36.

Goldberg, H., Sun, Y., Hickey, T.J., Shinn-Cunningham, B.G., & Sekuler, R. (2015). Policing fish at Boston's Museum of Science: Studying audiovisual interaction in the wild. *i-Perception*, *6*, 1–11.

Graham, C.H., & Kemp, E.H. (1938). Brightness discrimination as a function of the duration of the increment in intensity. *Journal of General Physiology*, *21*(5), 635–650.

Green, D.M., Mason, C.R., & Kidd, G.Jr. (1984). Profile analysis: Critical bands and duration. *Journal of the Acoustical Society of America*, *75*, 1163–1167.

Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *Journal of Acoustical Society of America*, *35*, 610.

Guttman, S.E., Gilroy, L.A., & Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychological Science*, *16*(3), 228–235.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(1), 1–13.

Harris, H., & Sagi, D. (2015). Effects of spatiotemporal consistencies on visual learning dynamics and transfer. *Vision Research*, *109*, 77–86.

Hussain, Z., McGraw, P.V., Sekuler, A.B., & Bennett, P.J. (2012). The rapid emergence of stimulus specific perceptual learning. *Frontiers in Psychology*, *3*, 226.

Hyun, J.-S., Woodman, G.F., Vogel, E.K., Hollingworth, A., & Luck, S.J. (2009). The comparison of visual working memory representations with perceptual inputs. *Journal of Experimental Psychology, Human Perception and Performance*, *35*, 1140–1160.

Julesz, B. (1962). Visual pattern discrimination. *Institute of Radio Engineers Transactions on Information Theory*, *8*, 84–92.

Julesz, B., & Guttman, N. (1963). Auditory memory. *Journal of the Acoustical Society of America*, 35, 895.

Keller, A.S., & Sekuler, R. (2015). Memory and learning with rapid audiovisual sequences. *Journal of Vision*, *15*, 1–18.

Kelly, D.H. (1961). Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements. *Journal of the Optical Society of America*, *51*(4), 422–429.

Kirchner, W.K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358.

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, *51*(13), 1457–1483.

Levi, D., & Saarinen, J. (2004). Perception of mirror symmetry in amblyopic vision. *Vision Research*, *44*(21), 2475–2482.

Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, *9*(12), 1–11.

McGeogh, J.A., & Irion, A.L. (1952). *The psychology of human learning*. New York: Longmans, Green.

Michalka, S.W., Kong, L., Rosen, M.L., Shinn-Cunningham, B.G., & Somers, D.C. (2015). Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks. *Neuron*, *87*(4), 882–892.

Mills, L., & Rollman, G.B. (1980). Hemispheric asymmetry for auditory perception of temporal order. *Neuropsychologia*, *18*(1), 41–48.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Noyce, A.L., Cestero, N., Shinn-Cunningham, B.G., & Somers, D.C. (2016). Short-term memory stores organized by information domain. *Attention, Perception & Psychophysics*, *78*(3), 960–970.

Oliveri, M., Koch, G., & Caltagirone, C. (2009). Spatial-temporal interactions in the human brain. *Experimental Brain Research*, *195*(4), 489–497.

Palmer, S.E., & Hemenway, K. (1978). Orientation and symmetry: effects of multiple, rotational, and near symmetries. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 691–702.

Piazza, E.A., Sweeny, T.D., Wessel, D., Silver, M.A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science*, *24*(8), 1389–1397.

Pollack, I. (1971). Depth of sequential auditory information processing: III. *Journal of the Acoustical Society of America*, *50*(2), 549–554.

Pollack, I. (1972). Memory for auditory waveform. *Journal of the Acoustical Society of America*, *52*(4), 1209–1215.

Pollack, I. (1973). Depth of visual information processing. *Acta Psychologica (Amsterdam)*, *37*(6), 375–392.

Pollack, I. (1990). Detection and discrimination thresholds for auditory periodicity. *Perception & Psychophysics*, *47*(2), 105–111.

Reardon, K.M., Kelly, J.G., & Matthews, N. (2009). Bilateral attentional advantage on elementary visual tasks. *Vision Research*, *49*(7), 691–701.

Rohenkohl, G., Gould, I.C., Pessoa, J., & Nobre, A.C. (2014). Combining spatial and temporal expectations to improve visual perception. *Journal of Vision*, *14*(4), 1–13.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, *51*(13), 1552–1566.

Saults, J.S., & Cowan, N.A. (2007). Central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology. General*, *136*(4), 663–684.

Sekuler, R. (1976). Seeing and the nick in time. In Siegel, M.H., & Zeigler, H.P. (Eds.) (pp. 176–197). New York: Harper & Row.

Sekuler, R., & Abrams, M. (1968). Visual sameness: a choice time analysis of pattern recognition processes. *Journal of Experimental Psychology*, *77*(2), 232–238.

Sekuler, R., Tynan, P., & Levinson, E. (1973). Visual temporal order: A new illusion. *Science*, *180*(4082), 210–212.

Serences, J.T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, *128*, 53–67.

Shipley, T.F., & Zacks, J.M. (2008). *Understanding events: From perception to action*. Oxford: Oxford University Press.

Sorkin, R.D. (1962). Extensions of the theory of signal detectability to matching procedures in psychoacoustics. *Journal of the Acoustical Society of America*, *34*, 1745–1751.

Umiltà, C., Stadler, M., & Trombini, G. (1973). The perception of temporal order and the degree of eccentricity of the stimuli in the visual field. *Studia Psychologica*, *15*, 130–139.

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences*, *1*(9), 346–352.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–438.

Watanabe, T., & Sasaki, Y. (2014). Perceptual learning: Towards a comprehensive theory. *Annual Review of Psychology*, *66*, 197–221.

Wilson, H.R. (1980). A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics*, *38*(3), 171–178.