



Memory and incidental learning for visual frozen noise sequences



Jason M. Gold^{a,*}, Avi Aizenman^b, Stephanie M. Bond^b, Robert Sekuler^b

^a Department of Psychological and Brain Sciences, Indiana University, United States

^b Volen Center for Complex Systems, Brandeis University, United States

ARTICLE INFO

Article history:

Received 25 April 2013

Received in revised form 7 September 2013

10 September 2013

Available online 25 September 2013

Keywords:

Incidental learning

Frozen noise

Memory

Reverse correlation

Mirror image

ABSTRACT

Five experiments explored short-term memory and incidental learning for random visual spatio-temporal sequences. In each experiment, human observers saw samples of 8 Hz temporally-modulated 1D or 2D contrast noise sequences whose members were either uncorrelated across an entire 1-s long stimulus sequence, or comprised two frozen noise sequences that repeated identically between a stimulus' first and second 500 ms halves ("Repeated" noise). Presented with randomly intermixed stimuli of both types, observers judged whether each sequence repeated or not. Additionally, a particular exemplar of Repeated noise (a frozen or "Fixed Repeated" noise) was interspersed multiple times within a block of trials. As previously shown with auditory frozen noise stimuli (Agus, Thorpe, & Pressnitzer, 2010) recognition performance (d') increased with successive presentations of a Fixed Repeated stimulus, and exceeded performance with regular Repeated noise. However, unlike the case with auditory stimuli, learning of random visual stimuli was slow and gradual, rather than fast and abrupt. Reverse correlation revealed that contrasts occupying particular temporal positions within a sequence had disproportionately heavy weight in observers' judgments. A subsequent experiment suggested that this result arose from observers' uncertainty about the temporal mid-point of the noise sequences. Additionally, discrimination performance fell dramatically when a sequence of contrast values was repeated, but in reverse ("mirror image") order. This poor performance with temporal mirror images is strikingly different from vision's exquisite sensitivity to spatial mirror images.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Humans are gifted pattern-recognizers, blessed with stunning ability to register, remember and exploit the similarities among sequences of sensory experiences. One especially useful approach to studying pattern recognition is to probe observers' ability to distinguish random stimulus sequences from random sequences onto which some form of structure has been imposed. Because random sequences comprise a homogenous pool of stimuli and can be devoid of semantic content, they put the research focus squarely on pattern-recognition's early stages – sensory processing and memory for features that are challenging to identify and process categorically (Kaernbach, 2004).

Over half a century, multiple researchers have exploited one simple but potentially informative strategy for imposing structure on random stimuli: repetition of a stored ("frozen") noise sample. Among the earliest uses of frozen noise, Guttman and Julesz (1963) showed that reiterating the same frozen auditory noise sample multiple times in succession generated characteristic auditory percepts, whose quality varied with the period of reiteration.

These observations were instrumental in Neisser's (1967) postulation of an echoic memory, a limited-duration auditory buffer. Later, Kaernbach (2004) showed that even just a single repetition of frozen noise could be discriminated from a non-frozen (that is, non-repeating) stimulus of equal duration. Recently, Agus, Thorpe, and Pressnitzer (2010) extended this work to explore the formation of auditory memory for sequences of random inputs. Their observers tried to discriminate between (i) 1-s long random sequences of auditory noise ('Random Noise'), and (ii) 1-s long sequences in which a single 500-ms auditory noise sequence was repeated so that it was presented twice in succession with no break between ('Repeated Noise'). Observers' performance discriminating between the two types of stimuli demonstrated their ability to exploit short term auditory memory – memory for the initial 500 ms of the stimulus – that had to be matched against the immediately ensuing 500 ms of the stimulus. Observers had good success in making this discrimination. Importantly, at random times during a block of trials, Agus et al. inserted a trial on which the very same Repeated Noise stimulus was recycled. At issue was whether experience cumulated over multiple trials with the same Repeated Noise stimulus would improve performance. Despite the many other stimuli intervening between successive presentations of a fixed Repeated Noise stimulus, performance

* Corresponding author.

E-mail address: jgold@indiana.edu (J.M. Gold).

with that stimulus did improve relative to randomly generated Repeated Noise stimuli. This improvement showed that observers not only formed short term memories, which allowed successive 500 ms sequences on a trial to be compared, but also were simultaneously forming longer-term memories, which cumulated over many trials. Moreover, Agus et al. showed that such learning of a fixed Repeated Noise exemplar was retained over experimental sessions, and was robust in the face of various acoustic transformations.

Agus et al. demonstrated the auditory system's remarkable ability to extract, store, and cumulate structure embedded in an arbitrary random sequence. But do these results reveal something specific to the processing of auditory information, or do they instead reflect some more generalized ability of human sensory systems to process arbitrary patterns, independent of the modality from which the patterns are received? Motivated by Agus et al.'s findings, and by the long-standing controversy about parallels between visual and auditory memory (e.g., Cohen, Horowitz, & Wolfe, 2009; Julesz & Hirsch, 1972; Visscher et al., 2007), we set out to examine vision's ability to support both short term memory and the longer term memory that Agus et al. demonstrated for audition. To do this, we adapted Agus et al.'s paradigm to explore observers' ability to discriminate and learn arbitrary visual noise sequences that are generated by temporally modulating stimulus contrast. We also applied reverse correlation analysis to the complex, temporally modulated stimuli used in our experiments, in order to identify in detail the strategies observers used when making their judgments (Neri & Heeger, 2002; Simoncelli, 2003).

2. Experiment one

Experiment 1 was modeled after the first of Agus, Thorpe, and Pressnitzer (2010)'s experiments, but used visual rather than auditory stimuli. In our experiment, observers tried to detect the presence (or absence) of a repeated sequence of visual contrast noise. Agus et al.'s noise stimuli were sampled and presented at 44 kHz, a value about twice the upper limit of hearing of otologically-normal young adults, but several log units above the temporal resolution of human vision. The many differences between the properties of vision and audition, including differences in temporal resolution, challenge attempts to make fair comparisons between the two (Visscher et al., 2007). In our experiments, we modulated the contrast of our visual stimuli across time as a step function at 8 Hz, a value near the peak of the human temporal contrast sensitivity function (Wilson, 1980). On each trial, the temporal modulation produced a sequence of eight items, each ~133 ms in duration.

On each trial, observers' task was to compare the sequence of the last four contrasts that they saw to the their memory of the sequence of the first four contrasts that they saw. We chose to use stimulus sequences whose units were four items in length because of evidence that visual short-term memory capacity has an upper limit of about four items (Phillips, 1974; Vogel, Woodman, & Luck, 2001). Observers were tested with two different kinds of visual noise: 1D noise, whose contrast was spatially uniform at any moment, but varied over time, at 8 Hz; and 2D noise, whose contrast varied in both time and space. As explained below, the 2D contrast variation in space produced a series of vertical stripes whose contrasts varied independently of one another over time.

2.1. Methods

2.1.1. Observers

Fourteen observers between the ages of 18 and 27 years participated in the experiment for a stipend of \$10 per experimental

session. All observers had normal or corrected to normal visual acuity, and were naive to the purposes of the experiment.

2.1.2. Apparatus

Unless otherwise specified, the following conditions were maintained across all experiments. Stimuli were presented against a uniform background of average luminance (19.03 cd/m^2) on a CRT monitor (Sony Trinitron UltraScan P780) at a resolution of 1024×768 pixels ($33 \times 24.5 \text{ cm}$) and refresh rate 75 Hz. Display contrasts were linearized by means of a calibration-based lookup table. Stimuli were generated and presented by an Apple iMac computer running Matlab (version 7.7) and extensions from the Psychophysics Toolbox (Brainard, 1997). Viewing was binocular through natural pupils. A viewing distance of 57 cm was enforced by means of a chin support. The computer display provided the only source of illumination in the room.

2.1.3. Stimuli

Gaussian white contrast noise was used to generate the contrast levels of all the stimuli in the experiment. Contrast was defined as $(L_{\text{pix}} - L_{\text{bg}})/L_{\text{bg}}$, where L_{pix} is the luminance of a given pixel, and L_{bg} is the background luminance (19.03 cd/m^2). Note that, according to this definition, contrast values could be either positive or negative. For our study, noise contrast levels were sampled from a normal distribution with mean equal to zero contrast and a variance equal to 0.2. Candidate samples more than ± 2 standard deviations from zero contrast were replaced by fresh samples, which restricted the range of contrast increments and decrements comprising any sequence. This algorithm for generating stimuli was intended to clamp the distinctiveness of individual sequences so that it would be difficult for observers to identify and explicitly recognize particular sequences.

Each stimulus sequence consisted of eight contrast levels presented in rapid succession to the same $4.1^\circ \times 4.1^\circ$ (128×128 pixels) region of the display. Each contrast level in an entire eight-item sequence was presented for 10 screen refreshes of the CRT display (~133 ms), which meant that a complete eight-item sequence played out in 1067 ms.

As mentioned earlier, the contrast noise was distributed spatially in two different ways within a 128×128 pixel stimulus square. To generate what we will call 1D noise, for each of the eight stimuli in a sequence, every pixel in a 128×128 pixel stimulus square was assigned the same contrast value. Thus, a 1D noise sequence consisted of a series of eight contrast values (3 bits of information). The other class of stimuli, which we will call 2D noise, was generated by assigning a different noise sample to each column of pixels in any square stimulus in an eight-item sequence. This produced, for each item in a sequence, 128 vertical stripes, each ~2 arcmin wide. The contrast levels of stripes within any item were independent of one another; moreover, the contrast levels varied independently over time, that is, across the eight items in a stimulus sequence. Thus, a 2D noise stimulus comprised a sequence of 128×8 contrast samples, or 10 bits of information, considerably more than in a sequence with our 1D stimuli.

For each kind of noise, 1D and 2D, we applied three different manipulations to the images' statistical structure over the eight items in a sequence (see Fig. 1). The manipulations produced three categories of stimuli, which we term Noise (N), Repeated Noise (RN), and Fixed Repeated Noise (FixRN). For stimuli of category N, the contrasts of the eight items comprising a sequence were independent of one another. In the case of repeated noise (RN), the first four frames (~533 ms) of the sequence repeated identically during the second half of the stimulus sequence, continuously and with no break in between halves. Finally, in the case of frozen or 'fixed' repeated noise (FixRN), a single randomly chosen RN stimulus was generated anew for each block of trials and was used

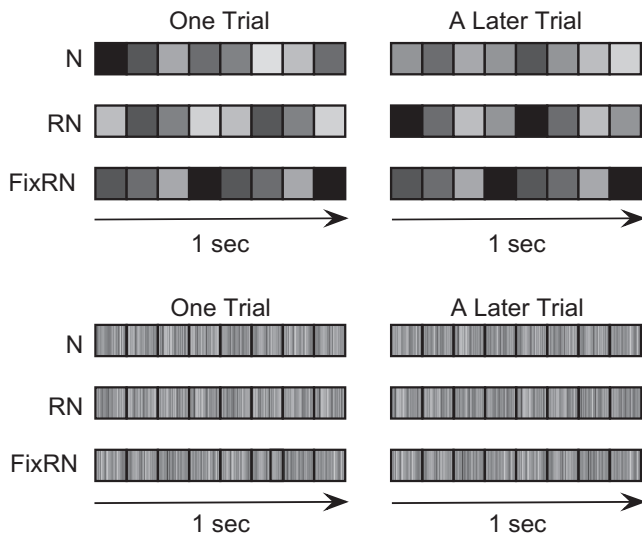


Fig. 1. Examples of Experiment 1's three kinds of stimuli (Noise (N), Repeated Noise (RN), and Fixed Repeated Noise (FixRN)). Note that the second half of an RN stimulus recycles the stimulus' first half; note also that an entire sequence FixRN repeats identically from one trial to a later trial. Top panel: 1D noise. Bottom panel: 2D noise.

throughout a block of trials whenever the experimental protocol called for a FixRN stimulus. Importantly, observers were not informed about the existence of a FixRN stimulus.

2.1.4. Design

All observers were tested with both 1D and 2D noise stimuli, with the order of testing counterbalanced across observers. Each observer completed four blocks of 200 trials per condition, making a total of 1600 trials. Each block contained 100 N trials, 50 RN trials and 50 FixRN trials. A unique FixRN stimulus was generated anew for each block and for each observer. The order of trials within a condition was randomly intermixed, with the constraint that two FixRN trials could not appear in immediate succession. Together with the relative proportions of the three stimulus categories, this constraint separated successive occurrences of the same FixRN stimulus by a mean of 3.9 trials ($SD = 3.0$) and a median of 2.4 trials.

2.1.5. Procedure

Before first being tested with either 1D or 2D stimuli, observers had an opportunity to practice the task for 20 trials in which N and RN stimuli of that type were presented. After these practice trials were completed and before actual testing began, observers completed an initial 3-min period of adaptation to the uniform background luminance of the CRT.

Stimuli on each trial were presented in the center of the display. Three hundred milliseconds after all eight items in a stimulus sequence had been presented, a message on the screen prompted the observer for a key press to signal whether the sequence was judged as 'Repeated' or 'Not repeated'. Accuracy feedback was given in the form of a high or low beep after each response. Observers were encouraged to rest after every 50 trials, but were asked to remain seated throughout the experiment.

2.2. Results

2.2.1. Overall performance

We first evaluated observers' overall task performance, computing d' for with repeated noise (RN) and fixed repeated noise (FixRN) stimuli. d' was calculated separately for each 200 trial block for each observer by taking the difference between $z(\text{hits})$ for either

RN or FixRN trials and $z(\text{false alarms})$ for N trials. A hit was defined as a response of "Repeated" to either an RN or FixRN stimulus; a false alarm was defined as a response of "Repeated" to an N stimulus. Fig. 2 shows the mean d' values for RN and FixRN trials, for both 1D (left panel) and 2D (right panel) noise.

As those panels of the figure show, for both 1D and 2D noise, overall performance was better for FixRN than for RN stimuli. The performance advantage enjoyed by FixRN stimuli over their RN counterparts show that observers were able to exploit the repetition of a particular FixRN sequence that was repeated at random times during a block of 200 trials. To assess the statistical reliability of this result, we carried out a two-factor, repeated-measures ANOVA whose variables were two types of repeated noise (RN and FixRN) and two levels of stimulus dimension (1D or 2D). The ANOVA confirmed that the effect of repeated noise type (RN vs. FixRN) was highly significant ($F(1,55) = 23.94, p < .0001$). Moreover, the dimensionality of the noise, 1D vs. 2D, significantly impacted performance, with 1D stimuli consistently producing better performance than 2D ($F(1,55) = 22.69, p < .0001$). There was no significant interaction between type of repeated noise and noise dimensionality type ($F(1,55) = 0.39, p = 0.53$).

2.2.2. Trial-wise performance

We further dissected observers' learning of FixRN stimuli by estimating the course of learning across trials within a block. Recall that each observer was tested with a unique sample of FixRN noise in each of 4 blocks of 200 trials, with each block containing 50 FixRN and 50 RN trials. To trace the average rate of learning within a block for both kinds of repeated noise, we computed performance across blocks and observers separately for individual trials within the 50 trial sequence for FixRN and RN stimuli. For example, performance for FixRN trial #1 was computed by combining all of the first FixRN trials across observers, regardless of whether the randomization within a block put that first FixRN stimulus first or somewhat later in the block. This produced a total of 56 trials (4 blocks per observer \times 14 observers), from which we computed the proportion of trials on which observer correctly classified the stimulus as FixRN (i.e., the proportion of 'hits'). Performance for the second occurrence of a FixRN stimulus computed by combining all of the second FixRN trials across all observers and computing the proportion of hits, etc. The same analysis was carried out for the RN trials.

Fig. 3 shows results of this trial-wise analysis. The circular symbols in each panel show percentage of hits for both RN trials (filled circles) and FixRN trials (open circles), with 1D noise plotted in the left panel and 2D noise plotted in the right panel. We also generated smoothed versions of these data by averaging percentage of hits within a 3-trial moving window (solid and dashed lines that roughly follow the data points). The width of this rectangular window is approximately equal to the median distance between successive FixRN trials. Error bars on the smoothed data were generated through bootstrap simulations, in which 200 simulated experiments were carried out by sampling the original data with replacement and computing the standard deviation across the resulting group of 200 simulated data sets (Efron & Tibshirani, 1993).

The data show a very gradual, improvement in performance over the trials within an block. Regression analysis of the raw data revealed a significant linear trend for 1D FixRN noise ($r = .57, p < .0001$), and a smaller, nearly significant linear trend for 2D FixRN noise ($r = .26, p = .07$). No such significant trends were seen for RN noise, either for its 1D form ($r = .19, p = .20$) or for its 2D form ($r = -.01, p = .98$).

2.2.3. Reverse correlation analysis

We next considered whether observers had adopted a consistent, distinct strategy as they tried to judge whether the second

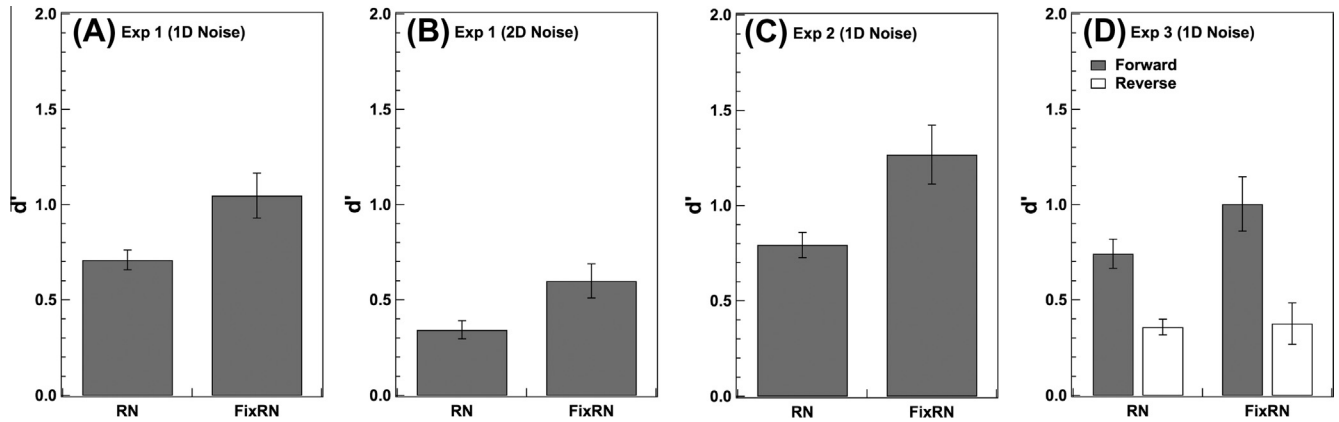


Fig. 2. Mean values of d' for Experiment 1 (panels A and B), Experiment 2 (panel C) and Experiment 3 (panel D). Panels A, C and D plot d' for 1D noise and Panel B plots d' for 2D noise. The hit rates for RN stimuli and fixRN stimuli were both referenced to false alarm rates for N stimuli. Error bars represent ± 1 s.e.m.

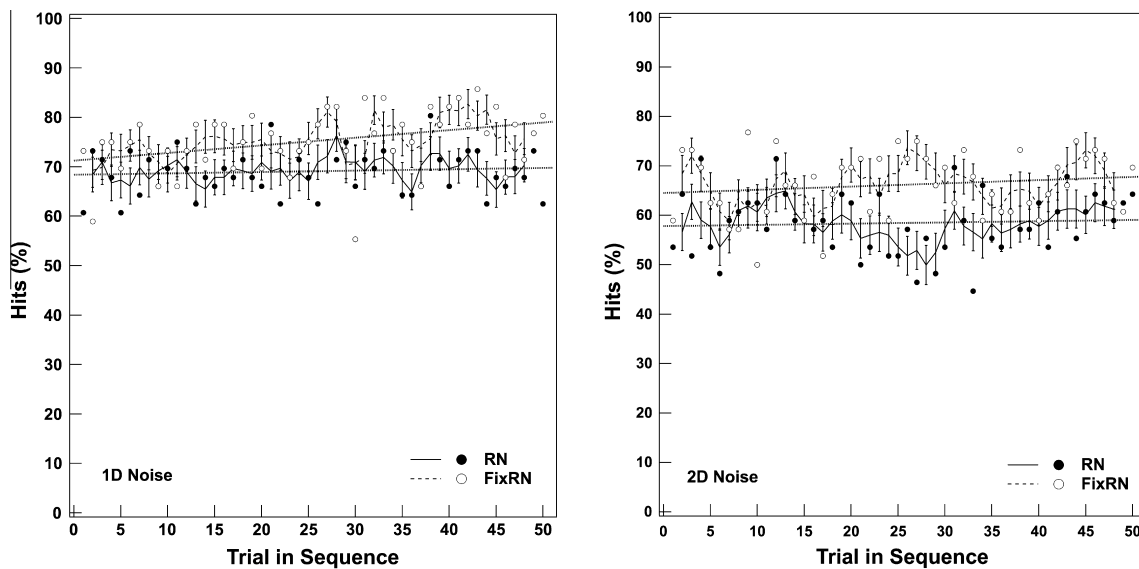


Fig. 3. Trial-wise performance in Experiment 1. In both figures, percentage of hits is plotted as a function of trial for both RN and FixRN stimuli. Left panel: 1D noise. Right panel: 2D noise. Lines roughly following the data points in each panel were produced by smoothing the raw data (circles) with a moving rectangular window three trials wide. Error bars represent ± 1 s.d. for the smoothed data, estimated by bootstrap simulations. Straight lines correspond to the best fitting (least squares) linear fit to each set of raw data.

half of a sequence did or did not repeat the first half. For example, did observers give equal weight to all the items in a sequence, or did they give preferential weight to a subset of items? To answer the question, we computed the correlation between the contrast of the noise in each of the eight items in a sequence and the observers' responses across trials. This technique is often called *reverse correlation* or *response classification* Marken and Sandusky (1975). Such an analysis typically requires a large number of trials to produce reliable correlations, and depends upon a variety of factors, including the magnitude of the observer's internal noise and the number of dimensions along which the stimulus varies (Murray, Bennett & Sekuler, 2002). In the case of our 2D noise stimuli, contrast varied along a relatively large number of dimensions (128 columns \times 8 stimulus items) relative to the total number of trials in our data set (8400 trials collapsed across observers and trial blocks). In the case of our 1D noise stimuli, contrast varied along far fewer dimensions (8 components per sequence), making the analysis of 1D noise much more suitable for reverse correlation. Thus, we restricted our analysis to the 1D noise condition.

In order to calculate the reverse correlation for the eight components in a 1D noise sequence, we regenerated vectors of the

eight contrast values that had been presented on all N and RN trials for all observers. These vectors were then sorted into four possible stimulus (s) – response (r) combinations (i.e., sNrN, sNrRN, sRNrN, sRNrRN). We then averaged these vectors within each s-r combination, and computed a *mean kernel* m as

$$m = (\mu_{sRNrRN} + \mu_{sNrRN}) - (\mu_{sRNrN} + \mu_{sNrN}) \quad (1)$$

We also computed the average squared deviation from the mean kernel, which is the *variance kernel* v :

$$v = (\sigma_{sRNrRN}^2 + \sigma_{sNrRN}^2) - (\sigma_{sRNrN}^2 + \sigma_{sNrN}^2) \quad (2)$$

The resulting mean and variance kernels are 8-element vectors, whose values represent the relative weights that observers placed on each of the eight components in a sequence. If there were no systematic relationship between the contrast of items occupying any of the eight ordinal positions in a sequence and observers' judgments, the resulting kernel weights would be zero for all eight positions. In the case of the mean kernel, values greater than 0 indicate a positive correlation between signed contrast and observers' tendency to respond that the noise was repeated; values less

than zero indicate a negative correlation between signed contrast and observers' tendency to respond that the noise was repeated. For example, a value greater than 0 at position i in the mean kernel indicates the observer was more likely to respond 'repeated' when the luminance of the noise at position i in the noise sequences was greater than the background and 'not repeated' when the luminance of the noise at position i in the noise sequences was less than the background. For cases where the value of the mean kernel is zero, the variance kernel has a relatively straightforward interpretation: namely, a value greater than zero in the variance kernel generally indicates high contrast in the noise at that point in the stimulus sequence was more likely to make observers respond 'repeated' than 'not repeated', regardless of its polarity; a value less than zero generally indicates high contrast in the noise was more likely to make observers respond 'not repeated' than 'repeated'. For cases where the values of the mean kernel are non-zero, the interpretation of the variance kernel is less clear, with departures from zero indicating the possible presence of either compressive or expansive nonlinearities within the system, depending upon the polarities of the kernels (Neri, 2010).

The results of this analysis are plotted in Fig. 4a. In this plot, each data point corresponds to the correlation estimate for an individual item in the sequence. The mean kernel is plotted in filled symbols, the variance kernel in open symbols. Error bars were estimated by bootstrap simulations, in which 2000 simulated experiments were carried out by sampling the complete original data set with replacement and computing the standard deviation across the resulting group of 2000 simulated kernels. We also ran an additional series of 2000 simulated experiments in which the noise samples shown on each trial were replaced by new random samples. These simulations allowed us to generate a ± 2 standard deviation confidence region around zero correlation. That is, values in this region fall within ± 2 standard deviations of what one would expect from a purely randomly generated kernel. The confidence region for the mean kernel is shown in light gray, the variance kernel in dark gray.

These data show something quite striking – that observers gave preferential weight to the fourth and, to a somewhat lesser extent, the eighth frame of a stimulus sequence. Note that these frames correspond to the final frames in each half of a repeating noise sequence. The fact that correlation is positive in the mean kernel for these items indicates that observers were more likely to respond 'repeated' when the luminance of the noise was brighter than the

background at these key frames (and likewise, more likely to respond 'not repeated' when the luminance of the noise was darker than the background). Interestingly, the variance kernel shows a negative peak at these key frames, indicating the possible presence of compressive nonlinearities in processing. However, a proper interpretation of the variance kernels would require the use of more sophisticated system identification tools, which in turn would require a much larger data set (Neri, 2010). Nevertheless, both the mean and variance kernels indicate the 4th and 8th frames of the stimulus had a particularly strong influence on observers' decisions in this task.

2.3. Discussion

Experiment 1 produced three major findings. First, as Fig. 2 showed, observers were able to learn the particular sequences (FixRN stimuli) of randomly generated temporal and spatiotemporal 1D and 2D contrast noise sequences, even though successive presentations of a FixRN stimulus was randomly interspersed amongst other stimuli that were drawn from the same distribution and generated in the same fashion. Second, performance with 1D sequences was significantly better than with 2D sequences, despite the fact that 2D sequences had many more samples that could potentially be compared across the two halves of a stimulus. Third, as Fig. 3 confirmed, a trial-wise analysis showed that learning took place gradually over a block of 200 trials – observers' performance significantly improved over the 50 repetitions of a FixRN stimulus. And fourth, reverse correlation analysis revealed that observers adopted the surprising strategy of preferentially weighting items at the two endpoints of a repeated noise sequence (Fig. 4a).

The results of Experiment 1 raise some interesting questions. First, the enhanced overall performance with a FixRN sequence relative to a set of RN sequences mirrors Agus et al.'s result of $\sim 3\times$ higher d' for FixRN relative to RN samples in the auditory domain. Our results show a less dramatic but certainly comparable effect ($\sim 1.7\times$ for 1D noise and $\sim 1.5\times$ for 2D noise). Despite this overall similarity, however, our trial-wise analysis suggests that the underlying learning process in our visual memory task is substantially different from what was reported for auditory memory. Whereas performance in our task improved gradually with repeated exposures to a fixed sequence of visual noise, Agus et al. found that performance improved sharply within just the first 10

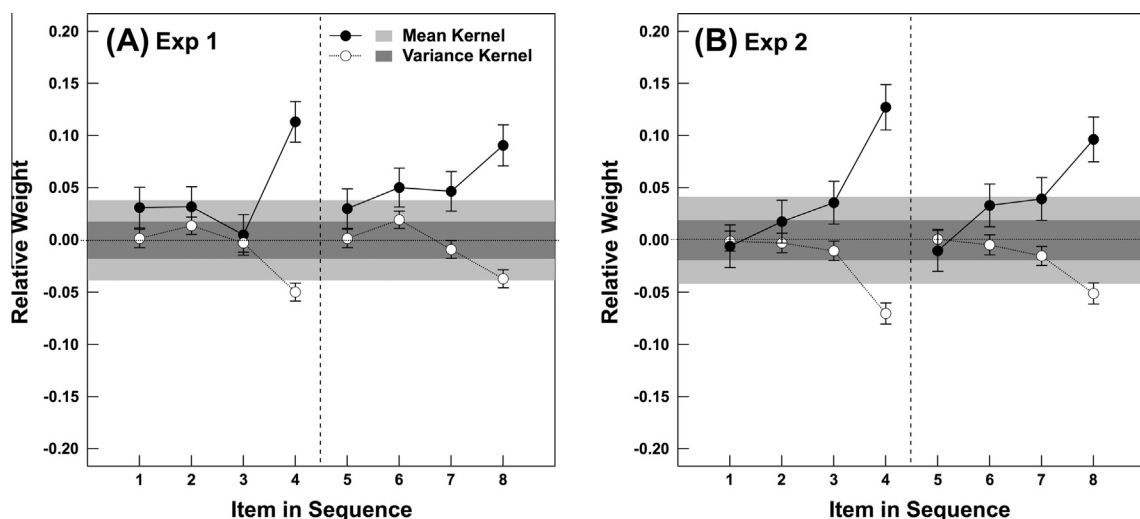


Fig. 4. Mean (closed symbols) and variance (open symbols) kernels estimated by reverse correlation for 1D noise stimuli in Experiments 1 and 2. Panel A: kernels for Experiment 1. Panel B: kernels for Experiment 2. Error bars on symbols represent ± 1 s.d., estimated by bootstrap simulations. Gray bands denote ± 2 s.d. confidence regions for each kernel type, estimated by bootstrap simulations.

exposures to a fixed repeating auditory noise sample, with little or no improvement thereafter. One possible reason for this difference between audition and vision may lie in the relatively low dimensionality of our noise samples (both 1D and 2D), compared to dimensionality of Agus et al.'s 44.1 kHz samples of auditory noise. Such low dimensional stimuli are far more likely to produce a heterogeneous range of individual FixRN sequences that spanned the range from particularly easy to particularly difficult to encode in memory. We address this possibility in Experiment 2, which tests observers with a common set of FixRN stimuli that yielded particularly good performance in Experiment 1.

Second, how robust is the visual system to various systematic transformations of the repeated portion of a noise sequence? One such transformation that has been explored extensively in the domain of spatial vision is mirror symmetry (e.g., Wagemans, 1995). In Experiment 3, we apply a mirror symmetrical transformation to the domain of temporal visual memory, testing whether observers can perform our repeated noise task with temporally mirror symmetric sequences.

Third, how durable is the learning that takes place for an individual FixRN sample in a single block of 200 trials? In Experiment 4, we take an initial step towards addressing this question, by testing observers with two identical trial-by-trial stimulus sequences occurring in immediate succession or separated by a 24 h period.

Fourth, the results of our reverse correlation analysis (Fig. 4a) raise the question of why observers preferentially weighted the fourth and, to a lesser degree, the eighth frames of a stimulus sequence? One possibility is that the profile revealed by the reverse correlation analysis simply reflects a natural, default strategy that observers automatically employ when searching for a temporally periodic pattern in a visual noise sequence. An alternative possibility is that observers chose this strategy to deal with the effects of intrinsic temporal uncertainty, which is a limiting factor in many temporally-based visual tasks (Cohn & Lasley, 1986). The presence of intrinsic temporal uncertainty would challenge an observer's ability to localize the temporal mid-point of a repeated noise sample, and therefore might lead observers to find and adopt a strategy that explicitly attempts to localize this critical point in the stimulus sequence. In Experiment 5, we test this possibility by inserting a variable duration temporal gap between the first and second halves of the stimulus sequence.

3. Experiment two

In Experiment 2 we were interested in exploring whether the difference between the gradual trial-by-trial improvements seen in Experiment 1 differed from the abrupt learning observed by Agus, Thorpe, and Pressnitzer (2010). This difference in learning may have been seen because the individual FixRN stimuli in Experiment 1 varied in the ease with which they could be encoded. To test this possibility, we first reanalyzed the 1D noise data from Experiment 1 by computing d' for each individual experimental block for each observer. The result of this analysis is shown as a scatterplot in Fig. 5a. These data confirm that there was in fact a wide range of d' values for the FixRN stimuli, and that this range was far wider than for the RN stimuli. Although individual observers' d' values for RN and FixRN sequences were significantly correlated ($r = .39, p < .01$), the fact that d' for FixRN varied much more than RN suggests that variation in FixRN performance may not have been entirely due to variation in observers' inherent sensitivities and the specific characteristics of the FixRN stimulus may have exerted some influence on performance. In order to identify what aspect of the FixRN stimuli led to better performance, we regenerated and sorted all of the FixRN exemplars used in Experiment 1 according to their corresponding d' (Fig. 5c). Close inspection of this figure

suggests that, in general, the contrast values of the FixRN exemplars on the higher end of the d' range tend to be more positive than negative in contrast, whereas the contrast values of the FixRN exemplars on the lower end of the d' distribution tend to have more of an even distribution of positive and negative contrast values. We tested this observation by computing the total summed contrast (i.e., the summed signed contrast values across frames) for each exemplar. These data are plotted in Fig. 5b, as a function of their corresponding d' . The closed symbols correspond to exemplars where d' was less than 1 (low d'), and the open symbols where d' was 1 or greater (high d'). These data show that the total summed contrasts of the exemplars for low d' s were distributed around a value that was shifted far lower than the exemplars for the high d' s. The average total summed contrasts for the low and high d' s were -0.23 and 1.19 , respectively. An independent samples t -test (two-tailed) showed that the difference between the mean low and high d' total summed contrast was highly significant ($t(54) = -3.72, p < .001$) and there was no significant difference between the variances of the two distributions ($F(29,25) = 1.01, p = .49$).

The outcome of these initial analyses led us to test whether the gradual trial-wise changes in performance seen in Experiment 1 were caused by the heterogeneity of its FixRN samples. If the gradual trial-wise changes in performance found in Experiment 1 resulted from variability in the 'goodness' of the FixRN samples, testing observers with only those FixRN samples that produced the best task performance in Experiment 1 could develop a clearer picture of the true underlying trial-by-trial learning effects. For this purpose, we decided to focus on a limited number of FixRN exemplars, testing a new set of observers' performance with a common set of FixRN stimuli, which produced high performance in Experiment 1. So, in each block of trials in Experiment 2, presentations of one of those select FixRN stimuli were randomly intermixed with newly-generated N and RN stimuli.

3.1. Methods

3.1.1. Observers

Twelve observers over the age of 18 years participated; none had served in Experiment 1. All had normal or corrected to normal visual acuity and were naive to the purposes of the experiment.

3.1.2. Stimuli, procedure and design

The stimuli and procedure were identical to those described in Experiment 1 for the 1D noise condition, with one exception. Rather than randomly generating a novel FixRN stimulus for each experimental block and observer, FixRN stimuli were restricted to the four FixRN stimuli that had yielded the highest overall performance in Experiment 1 (i.e., the four sequences shown at the rightmost end of Fig. 5c). All observers were tested with the same four FixRN stimuli, but in different orders of blocks, which were randomized for each observer according to a 4×4 Latin square (replicated 3 times to accommodate all 12 observers).

3.2. Results

Fig. 2c presents the overall mean d' values for 1D RN and 1D FixRN stimuli. These values were computed as in Experiment 1, with hit rates for RN or FixRN stimuli each evaluated relative to the false alarm rates for the intermingled N stimuli. Note the strong similarity of these results to ones obtained with 1D stimuli in Experiment 1 (Fig. 2a). Although the absolute d' values were slightly higher for both RN and FixRN than in Experiment 1, the ratio of d' values for FixRN and RN was nearly identical (a factor of ~ 1.6). Just as in the previous experiment, the difference between FixRN and RN d' values in Experiment 2 was highly significant ($t(47) = 3.33, p < .001$). Thus, the overall learning effect seen in Experiment 1 was

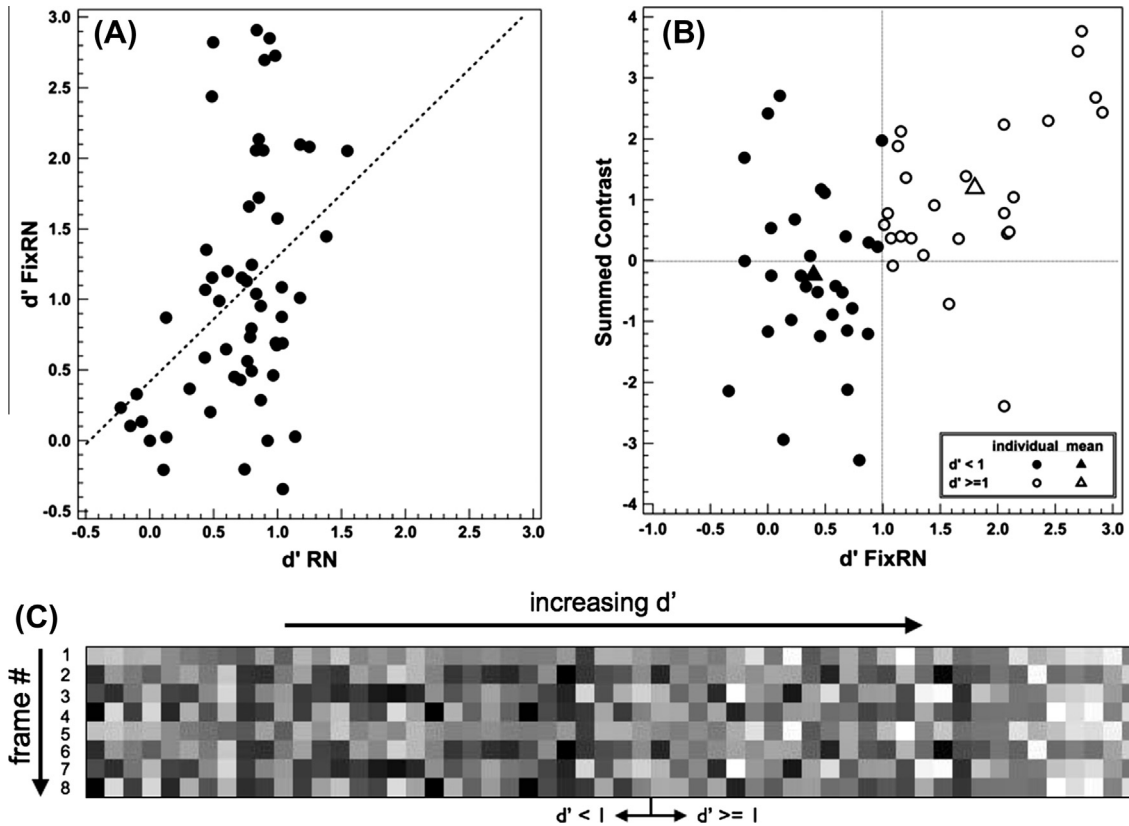


Fig. 5. (A) d' for the FixRN stimuli for individual observers in Experiment 1, plotted as a function of d' for RN stimuli; (B) total summed contrast (signed) for the FixRN stimuli shown to individual observers in Experiment 1, plotted as a function of d' ; (C) the FixRN stimuli shown to observers in Experiment 1, sorted in order of d' . Each pixel corresponds to a single frame, with time progressing from top to bottom.

replicated when all observers were tested with the same four FixRN sequences that previously promoted particularly good FixRN performance.

Fig. 6a plots the mean trial-wise performance for FixRN and RN sequences. The height and slopes of the two curves are similar to those obtained with 1D noise in Experiment 1 (lefthand panel of Fig. 3). Fig. 4b plots the results of a reverse correlation analysis, performed in the same fashion as described in Experiment 1. As before, the reverse correlation analysis suggests that observers assigned disproportionate weight to items that occupied the 4th and 8th ordinal positions in the stimulus sequence.

3.3. Discussion

Taken together, the results of Experiment 2 indicate that two of the effects found in Experiment 1 – the gradual rather than abrupt trial-wise learning, and the disproportionate weighting of the 4th and 8th stimulus items in a sequence – were not due to the use of a large, heterogenous sample of FixRN stimuli. Similar results were obtained in both experiments, even though in Experiment 1 observers were given unique FixRN samples for each block of trials, while in Experiment 2 a common set of four FixRN samples were recycled across all observers.

One somewhat puzzling aspect of these data that is worth pointing out is the fact that average performance with the FixRN stimuli was nearly identical to what was obtained in Experiment 1, despite the fact that we used a restricted set of FixRN sequences, all of which yielded exceptionally high performance in Experiment 1. One possible explanation this surprising result is that the high level of performance obtained in Experiment 1 with these four FixRN

exemplars resulted not solely from some special qualities of the exemplars themselves, but also from the extraordinary performance of observers in Experiment 1 who happened to be assigned those particular exemplars. We looked more closely at this possibility by plotting the sensitivities of individual observers for each of the four FixRN samples used in Experiment 2 (Fig. 6b). In this graph, individual observer's data are represented as numerals ranging from 1 to 12. Results from each of the four FixRN stimuli are plotted in a separate column. The average sensitivity across observers for each FixRN sample is also shown in each column, by the filled symbols. These data show that there was substantial variability among mean d' values associated with different FixRN sequence, and that there was also substantial variability in observers' sensitivity within each FixRN sequence. Also note that each individual observer's sensitivities tended to be consistent with respect to the sensitivities of the other observers. For example, observer 10 consistently had the highest sensitivities across all of the FixRN samples, whereas observers 5 and 7 tended to have the lowest sensitivities. These patterns reveal that there were both stimulus-based and observer-based sources of variability present in our data. They also strongly suggest that the exceptional performance found with these four FixRN samples in Experiment 1 was probably not caused by the particular qualities of the FixRN samples, but rather resulted from individual differences in observer sensitivities.

Thus, the results of Experiment 2 rule out the possibility that the trial-wise learning and disproportionate weighting of particular items in a sequence, which were seen in Experiment 1, were caused by that experiment's large, heterogenous set of FixRN samples. We revisit one of these effects (the disproportionate weighting of the 4th and 8th stimulus frames) in Experiment 5.

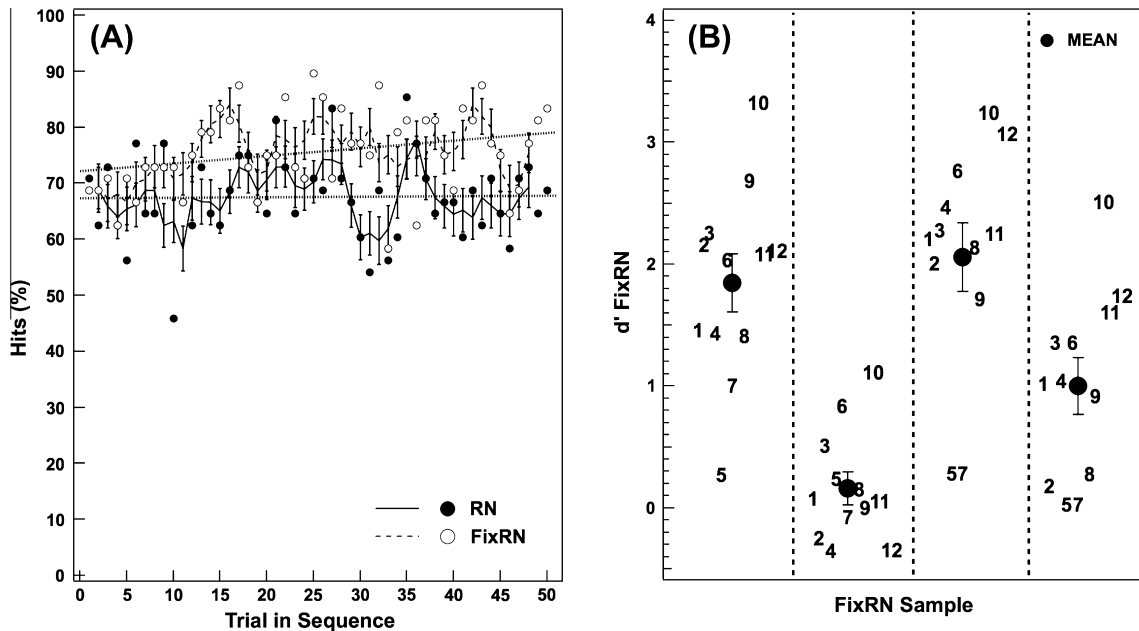


Fig. 6. (A) Trial-wise performance in Experiment 2. Percentage of hits is plotted as a function of trial for both RN and FixRN stimuli. Lines roughly following the data points were produced by smoothing the raw data (circles) with a three-trial-wide roving window. Error bars represent ± 1 s.d. for the smoothed data, estimated by bootstrap simulations. Straight lines correspond to the best fitting (least squares) linear fit to each set of raw data. (B) Individual observer d' s for each of the four FixRN exemplars used in Experiment 2. Symbol numbers correspond to individual observers. Filled symbols correspond to the mean across observers for each FixRN sample. Error bars represent ± 1 s.e.m.

4. Experiment three

In the preceding two experiments, the notion of ‘repetition’ was used in the strictest possible sense – an exact replication of items in the first half of a stimulus sequence, including of course, the order in which items appeared in the two halves. However, a similar manipulation to which the visual system is thought to be particularly sensitive with respect to spatial patterns is a *reverse* replication of the first half of a stimulus – i.e., mirror symmetry (Barlow & Reeves, 1979; Bruce & Morgan, 1975; Jenkins, 1982). Some of the most recent evidence for the sensitivity of the visual system to spatial mirror symmetry comes from neuroimaging research demonstrating that the dorsolateral extrastriate visual cortex contains symmetry-sensitive neuronal representations (Cattaneo et al., 2011). Furthermore, overt visual attention and human eye fixations are attracted by local, mirror symmetry in spatial patterns (Kootstra, de Boer, & Schomaker, 2011).

Despite many studies of the visual system’s striking sensitivity to spatial mirror symmetry, remarkably little is known about the complementary case, sensitivity to mirror symmetry in the time domain. Therefore, Experiment 3 explored the sensitivity of the human visual system to temporal mirror symmetric transformations. Observers performed the same noise-repetition discrimination task as in Experiments 1 and 2, but with sequences that contained either an exact repetition of individual items’ order, or a temporally-mirror symmetric repetition in the second half of the stimulus. If the visual system’s sensitivity to temporal and spatial mirror symmetry are comparable, we would expect performance with a temporally symmetric pattern to match or perhaps even exceed that of an untransformed temporal repetition.

4.1. Methods

4.1.1. Observers

Twenty-one observers over the age of 18 participated in the experiment. All had normal or corrected to normal visual acuity and were naive to the purposes of the experiment.

4.1.2. Stimuli, procedure and design

With two exceptions, the stimuli, procedure and design were identical to those used with 1D noise in Experiment 1. First, this experiment introduced two different modes of sequence repetition: a Forward condition, in which the second half of a repeated stimulus sequence was an exact temporal duplicate of the first half (i.e., a direct replication of the 1D condition of Experiment 1); and a Reverse condition, in which the second half of a repeated stimulus sequence was a temporally reversed version of the first half. In the Reverse condition, the initial frame of the second half of the stimulus was removed, in order to avoid having the middle two frames be identical (and thus appear as a single frame that was twice the duration of the other frames). As a result, the Reverse condition contained only seven unique frames, rather than eight. Second, a between-subjects design was used, with 10 observers being randomly assigned to the Forward condition and 11 to the Reverse condition. We chose a between-subjects design because of an issue that arose during pilot tests using a within-subject design in which Forward and Reverse conditions were intermingled. When observers could not know ahead of time whether, from trial to trial, a repeat would be either Forward or Reverse, overall performance was quite poor.

4.2. Results and discussion

Figs. 2d, 7 and 8 show the results of Experiment 3. Fig. 2d plots mean d' values for RN and FixRN noise, separating performance in the Forward and Reverse conditions. These data reveal two interesting effects. First, observers had a difficult time detecting a noise repetition when it was presented as temporally mirror symmetric. This is demonstrated by the lower overall d' s for both RN and FixRN noise samples. Second, unlike in the Forward condition, the lack of difference between RN and FixRN in the Reverse condition suggests that observers did not learn and utilize the characteristics of the mirror symmetric FixRN noise samples over the course of a single block. An ANOVA with repeated measures on one factor (noise repetition type) confirmed that performance in the Forward

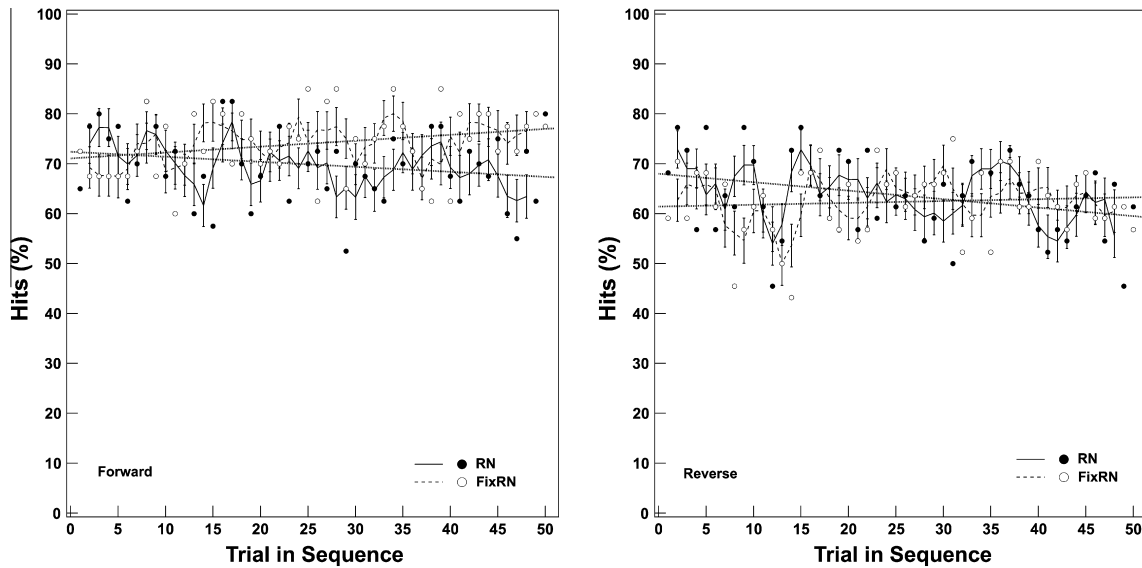


Fig. 7. Trial-wise performance in Experiment 3, shown as percentage of hits plotted against successive trials for both RN and FixRN stimuli. Results in the lefthand panel are for trials on which a sequence was repeated in Forward direction (as in Experiments 1 and 2); results in the righthand panel are for trials on which a sequence was repeated in Reverse order. Lines roughly following the data points were produced by smoothing the raw data (circles) with a three-trial-wide roving window. Straight lines correspond to the best-fitting (least-squares) linear fits to the raw data. Error bars represent ± 1 s.d. for the smoothed data, estimated by bootstrap simulations.

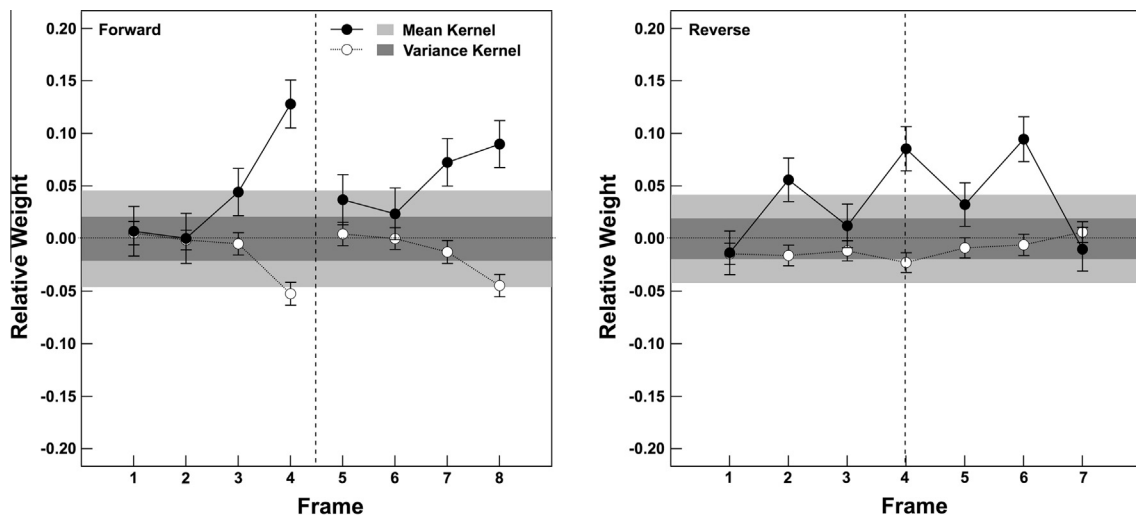


Fig. 8. Mean (closed symbols) and variance (open symbols) kernels estimated by reverse correlation for the two types of 1D noise stimuli used in Experiment 3. Left panel: kernel for stimulus sequences whose first four items repeated in the same direction (forward) as originally presented. Right panel: kernels for stimulus sequences whose first four items were repeated in reverse order. Error bars on symbols represent ± 1 s.d., estimated by bootstrap simulations. Gray bands denote ± 2 s.d. confidence regions for each kernel type, estimated by bootstrap simulations.

condition was significantly better than in the Reverse condition ($F(1,83) = 10.71, p < .0001$). The overall difference between RN and FixRN noise was only marginally significant ($F(1,82) = 3.5, p = .06$), as was the interaction between noise repetition type and stimulus direction ($F(1,82) = 2.86, p < .09$).

The results of a trial-wise analysis are shown in Fig. 7. The left and right panels of that figure show the results for the Forward and Reverse conditions, respectively. Trial-wise performance in the Forward condition replicated the results found in the previous experiments, showing a very gradual increase in performance across trials with the FixRN stimuli. However, trial-wise performance with the mirror reversed stimuli was markedly different, showing a gradual decrease in performance across trials for the FixRN stimuli. That is, performance seems to decline over time as observers are exposed to additional samples of the FixRN stimulus.

Regression analyses of the raw data revealed only marginally significant linear trends in all conditions: Reverse RN noise: $r = -.31, p = .07$; Reverse FixRN noise: $r = .08, p = .07$; Forward RN noise: $r = -.20, p = .07$; Forward FixRN noise: $r = .26, p = .06$.

Results of reverse correlation analyses are shown in Fig. 8. The left and right panels plot the weights for the Forward and Reverse conditions, respectively. The pattern of weights across items in the Forward condition closely replicated the characteristic pattern of preferential weighting for the 4th and 8th frame found in previous experiments (see Fig. 4). However, a very different weighting pattern emerged in the Reverse condition. For Reverse stimuli, the mean kernel shows that observers appeared to be ‘looking’ for a temporal contrast pattern that passed through approximately two periodic cycles of contrast variation across the duration of a stimulus sequence. One possible reason for the surprising result

that observers' performance with the FixRN stimulus declined across trials is that observers may have become progressively more inclined to adopt the above periodic strategy over the course of an experimental block. Unlike the strategy used by observers in the Forward condition – which likely helped observers mitigate the negative effects of temporal uncertainty (a possibility explored in more detail in Experiment 5) – the pattern of weights adopted in the Reverse condition may not have provided a similar benefit to outweigh the costs associated with applying a fixed periodic strategy.

5. Experiment four

The trial-wise analyses of hits in Figs. 3, 6 and 7's show that observers can learn the characteristics of an arbitrary noise sequence over the course of multiple exposures to a FixRN sequence. This confirms the remarkable encoding ability of short term visual memory. But what are the longer-term properties of memories that are formed? Are they simply reset after a brief 200-trial session is complete, when a new FixRN sequence is introduced? Or are they perhaps stored for a longer period of time?

Agus, Thorpe, and Pressnitzer (2010) provided evidence of some longer-term retention of specific noise samples in memory. One of their experiments presented observers with the same FixRN sample in two successive blocks, with the blocks separated by an average of 16.5 days. Although they found no overall improvements in performance across successive days (i.e., d' for FixRN did not improve from day 1 to day 2), they did find differences in the rate of initial trial-wise learning that took place for the FixRN stimulus for each block. Specifically, on day 1 they observed their typical pattern: a rapid increase in performance across the first 5–10 trials, with asymptotic performance thereafter. On day 2, they found that the initial ramping up of performance during the first 5–10 trials was virtually absent, with performance becoming asymptotic at or near the block's start. This result suggests that observers did not have to relearn the specific noise patterns to which they had been exposed several weeks earlier, presumably because they had stored them in long-term memory.

Experiment 4 was designed to determine whether with random visual sequences learning was as robust to the passage of time as was Agus et al. found with auditory sequences. To do this, we tested each observer in two sessions using exactly in both sessions the same stimuli and trial sequences. The sessions were distributed over two successive days, separated by at least 24 h. We chose this separation between sessions rather than the much longer one used by Agus et al. in order to increase the chances that we would see significant retention of visual memory for learned FixRN stimuli across days.

In addition to examining the robustness of memory for specific noise samples, we were also interested in revisiting the effects of the dimensionality of our noise stimuli. We speculated that noise patterns that only vary in the temporal domain may have a very limited set of memorable 'features', and this may be partially responsible for our inability to replicate the trial-wise learning effects found by Agus et al.. Although the 2D noise that we tested observers with in Experiment 1 contained many 'features' more than did the 1D noise (128 spatial \times 8 temporal samples of 2D noise vs. just 8 temporal samples of 1D noise), it might be that each 2D spatial sample's small retinal angle (~ 1 minarc) promoted spatial mismatches or spatial smearing across successive temporal samples. Such spatial mismatches, if they occurred on RN or FixRN trials and if they were not well-correlated between items in a sequence's two halves, would have disrupted perception of the correspondence between samples in those half-sequences. In order to examine this possibility, we tested observers with lower resolution

2D conditions. Specifically, we generated 2D noise stimuli that had only two spatial noise samples per frame, rather than the full resolution versions used in Experiment 1.

5.1. Methods

5.1.1. Observers

Sixteen new observers were recruited from same age group as in the previous experiments. All had normal or corrected to normal visual acuity and were naive to the purposes of the experiment.

5.1.2. Stimuli, procedure and design

The stimuli and procedure were similar to those described in the Experiment 1. Observers were tested with both 1D and low-resolution 2D noise. The 1D noise was identical what was used in the previous experiments. The lower resolution 2D noise contained only two spatial samples of noise on each frame instead of 128, and each half of the 128×128 spatial region on a given frame was uniformly assigned one of these two samples. This appeared as two large vertical bars, each 64×128 pixels (128×256 arcmin) that defined the left and right halves of the entire 128×128 pixel stimulus region. New samples of spatial noise were generated for each temporal frame of the stimulus (i.e., 2 spatial samples \times 8 temporal frames).

Each observer was tested with three unique FixRN samples, one in each of three blocks of 200 trials (50 FixRN, 50 RN trials, and 100 N trials). All observers participated in both the 1D and 2D noise conditions, and were retested with the exact same stimuli (literally, identical trial-by-trial) 24 h later. The order of the noise conditions was counterbalanced across observers, so that half of the observers saw the 1D noise condition blocks first and the other half saw the 2D noise condition blocks first. This order was then replicated during the second day of testing.

5.2. Results and discussion

The results of Experiment 4 are shown in Figs. 9–11. Fig. 9 shows average d' for RN and FixRN noise for on each day of testing for 1D and 2D stimuli. A $2 \times 2 \times 2$ repeated measures ANOVA tested the effects of each of these variables and their interactions. There were significant main effects for all three variables: performance with FixRN was better than RN ($F(1,47) = 67.3, p < .0001$), performance on day 2 was better than on day 1 ($F(1,47) = 24.9, p < .0001$), and performance was better with 2D than 1D noise ($F(1,47) = 4.88, p < .05$). There was also a significant 2-way interaction between repeated noise type and day of testing ($F(1,47) = 17.9, p < .0001$), with FixRN noise showing greater improvement across days than RN. Finally, there was also a significant three-way interaction ($F(1,47) = 5.71, p < .05$). The improvements in performance from day 1 to day 2 for FixRN but not RN stimuli are consistent with the idea that observers retained some memory of the specific FixRN samples they saw on day 1, and were able to use this memory to enhance performance on day 2.

The results of the trial-wise analysis are shown in Fig. 10. The top two panels plot performance for 1D noise and the bottom two panels 2D noise. The left two panels correspond to day 1 and the right two panels day 2. For the 1D noise on day 1, trial-wise performance was similar to what was seen in previous experiments – a gradual increase in performance over time. On day 2, performance simply maintained a similar level across time. However, the pattern was quite different for 2D noise. Here, for the first time in this series of experiments, we saw a rapid improvement in performance for FixRN stimuli on day 1 across the first 10 trials, during which, on average, just three samples of a FixRN stimulus would have been presented. We verified the reliability of this result by fitting a line to the first ten trials for 2D FixRN stimuli on

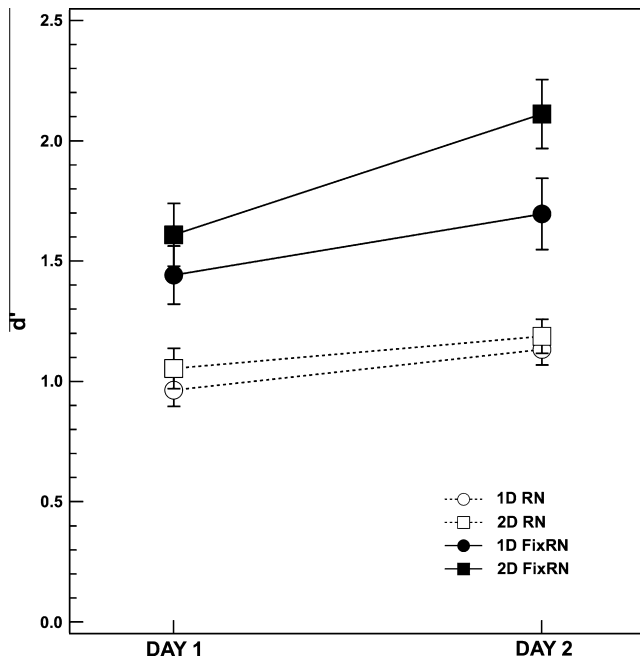


Fig. 9. Values of d' produced by 1D and 2D versions of RN and FixRN stimuli on the first and second days of testing.

Day 1, finding the slope of that line, and then using a randomization test¹ to determine how often the slope of the hit rates of any 10 trials chosen randomly without replacement from all trials would exceed the slope found for hit rates over the first 10 actual trials. For lines that were fit to the 380,000 permuted hit rates, less than one permutation in 10,000 met the criterion ($p < .0001$), supporting the suggestion that learning over the first ten trials with 2D FixRN is indeed atypically strong. To test whether this result was some artifact of the smoothing process, we performed the analogous randomization tests on the raw, non-smoothed data. Similar to what we found with the smoothed data, fewer than one per thousand slopes fit to randomized data equaled or exceeded the slope of the line fit to the actual results from the first ten trials with 2D FixRN stimuli.

These results, with smoothed and non-smoothed data, resemble the rapid learning Agus, Thorpe, and Pressnitzer (2010) found with random samples of auditory noise. Further, performance on day 2 in our experiment fails to show such a trend. Instead, performance begins at a level that is already higher than that at the end of the first day's performance, and this level is maintained across the rest of the session. Finally, we also see a progressive widening of the separation between FixRN and RN performance over time with 2D, but not 1D stimuli. Apparently, doubling the amount of spatial information available in each item made it possible for observers to rapidly learn the characteristics of a FixRN noise sample – something they were unable to do with just a single temporally varying sample. However, it is unclear whether this increase in the rate of learning resulted from a doubling of the amount of *spatial* information available to observers or was simply due to an increase in the total number of features present in the stimulus. Future experiments would need to be carried out to distinguish between these two possibilities.

Fig. 11 shows the results of reverse correlations on data from Experiment 4. For the sake of simplicity, only the results for the

mean kernels are shown in these figures. In the leftmost panel, weights for 1D noise are plotted separately for each day of testing; for 2D noise, the weights for a stimulus' left and right sides are plotted in separate panels, with results for the stimulus' left-hand half shown in the middle panel and results for the stimulus' right-hand half in the rightmost panel. These data show that, despite the significant learning effects seen in overall performance, the strategy employed by observers did not change in any obvious way from day 1 to day 2. Further, observers' strategies did not differ in any obvious way between either 1D and 2D noise or the left and right sides of the 2D stimulus. In most conditions, we see the characteristic preferential weighting of the middle and end frames of the stimulus sequence.

6. Experiment five

Our final experiment tested a possible explanation for why observers in our previous experiments seemed to give disproportionately heavy weight to the 4th and 8th items in stimulus sequences. We hypothesized that that this disproportionate weighting might reflect observers' intrinsic uncertainty about the point in time at which the first half of a repeated sample of noise ended and the second half began. That is, observers might disproportionately weight the 4th frame in the sequence because they are trying to anticipate when the temporal mid-point of the stimulus sequence would occur, at which time they could start the process of comparing the first and second halves of the sequence. Further, because they would have placed a disproportionate amount of weight on the 4th frame in the sequence, they would also be forced to place a disproportionate amount of weight on the corresponding 8th frame of the sequence, in order to determine whether or not the heavily-weighted 4th frame was repeated.

If temporal uncertainty had encouraged observers' uneven weighting strategy in the previous experiments, reducing that uncertainty should promote a weighting strategy that is more evenly distributed across a stimulus sequence. One possible approach to reducing temporal uncertainty would be to delineate a sequence's two halves by inserting a temporal boundary marker between those halves. To test this hypothesis, we inserted brief temporal gaps of varying duration between the first and second halves of each stimulus sequence. The temporal gap should provide a clear temporal marker for the mid-point of the stimulus sequence, which should in turn reduce observers' temporal uncertainty about when the sequence's mid-point occurs.

6.1. Methods

6.1.1. Observers

Twenty-four naive observers were recruited for Experiment 5. One observer did not complete the experiment, and thus was excluded from the data analysis. All observers had normal or corrected to normal visual acuity.

6.1.2. Stimuli, procedure and design

The stimuli and procedure were identical to the 1D noise condition from Experiment 1, with just one exception. A temporal gap of zero contrast was introduced between the 4th and 5th items in each stimulus sequence. Note that, with zero contrast, the entire screen was at mean luminance. Three different temporal gap durations were tested: 0, 133 and 400 ms. The 0 ms gap duration replicated the conditions of the previous experiments, that is, no gap was inserted; the 133 and 400 ms gap durations corresponded to the durations of a single item in a sequence, and three items in a sequence, respectively. For each of the three gap durations, an observer was tested in two blocks of 200 trials. The order of

¹ Because there are more than 6.5 billion ways to choose 10 items without replacement from a set of 48 items, it would have been impractical to do permutation tests on our data. Therefore we opted for randomization tests, using a large number of randomizations (380,000) in each test.

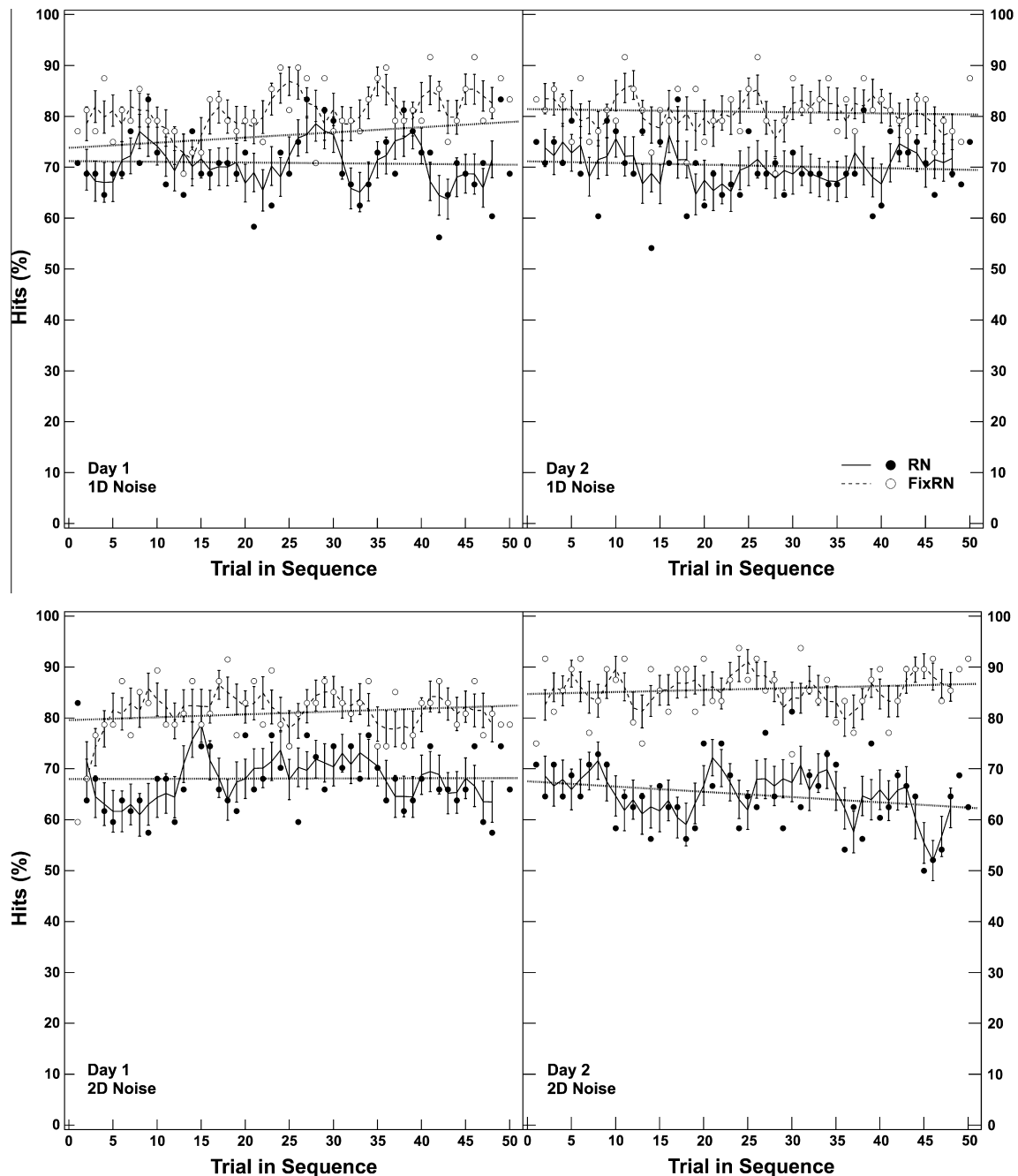


Fig. 10. Trial-wise performance in Experiment 4, shown as percentage of hits plotted as a function of successive trials for RN (filled symbols) and FixRN stimuli (open symbols). Panels in the upper row show results for 1D stimuli; panels in the lower row show results for 2D stimuli. The left-hand panels show results from Day 1 of testing; the right-hand panels show results from Day 2 of testing. Smoothing the raw data with a three-trial-wide roving window produced the lines shown connecting data points. Error bars represent ± 1 s.d. for the smoothed data, estimated by bootstrap simulations. Straight lines correspond to the best fitting (least squares) linear fit to each set of raw data.

conditions was counterbalanced across observers, as well as across the three gap duration conditions.

6.2. Results and discussion

Figs. 12 and 13 show results of this experiment. Fig. 12 presents d' for RN and FixRN noise as a function of gap duration. Most importantly, the data show that, for both RN and FixRN stimuli, d' increased monotonically with the duration of the gap separating the first half of a stimulus from its second half. This suggests that the presence of a temporal gap between noise repetitions did succeed in reducing temporal uncertainty. Additionally, as found in

the previous two experiments, d' was higher for FixRN than RN stimuli, showing that observers were able to learn the fixed repeated sample of noise. Finally, the effect of gap duration did not differ between RN and FixedRN stimuli. A repeated measures ANOVA, with two types of repeated stimuli and three gap durations confirmed a highly significant effect of both repeated noise type ($F(1,45) = 19.8$, $p < .0001$) and gap duration ($F(2,90) = 17.22$, $p < .0001$), but no significant interaction between repeated noise type and gap width ($F(2,90) = .17$, $p = .83$).

Most directly relevant to our hypothesis about temporal uncertainty are the results of the reverse correlation analysis, which are plotted separately for each gap duration in the top left panel of

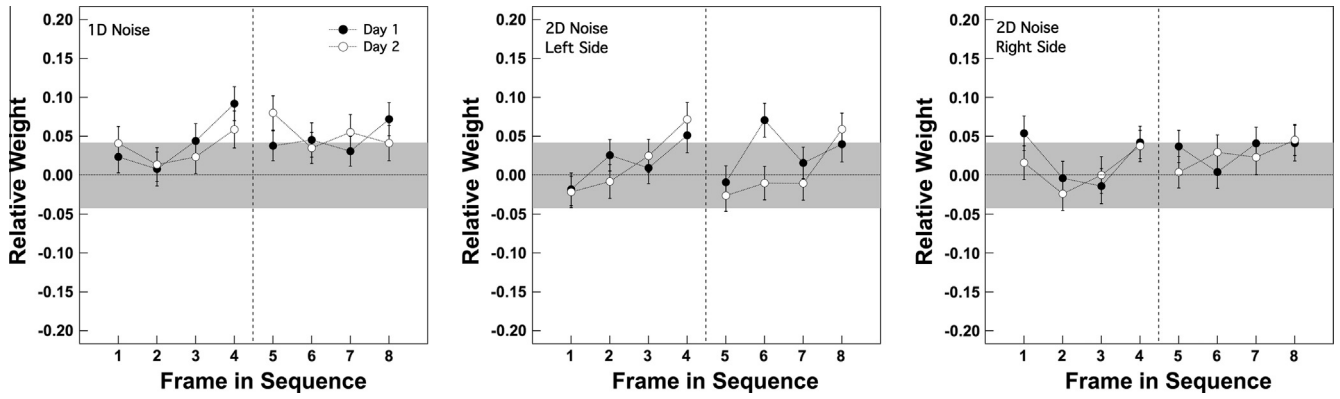


Fig. 11. Mean kernels estimated by reverse correlation for the 1D and 2D noise stimuli used in Experiment 4. Values derived from an analysis of the first day’s testing are shown as filled symbols; values from the second day’s testing are shown as open symbols. Left panel: mean kernel for 1D stimulus sequences. Middle panel: mean kernel for contrasts comprising the left half of a 2D stimulus sequence; right panel: mean kernel for contrasts comprising the right half of a 2D stimulus sequence. Error bars represent ± 1 s.d., estimated by bootstrap simulations. Gray band in each panel denotes ± 2 s.d. confidence region, estimated by bootstrap simulations.

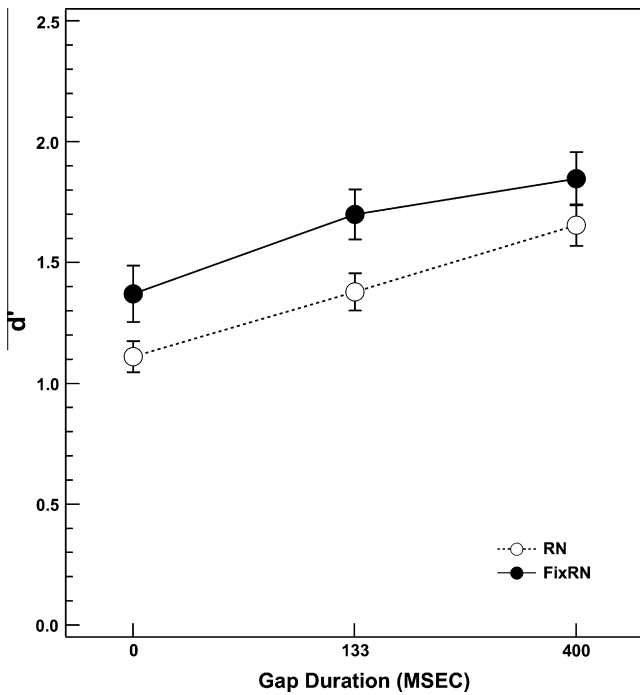


Fig. 12. d' values associated with gaps of different duration inserted into sequences between the fourth and fifth items. Open symbols show results for RN stimuli; closed symbols are for FixRN stimuli.

Fig. 13. As with Experiment 4, for the sake of simplicity only results for mean kernels are shown in these figures. Recall that our hypothesis was that the presence of a temporal gap between the first and second repeated noise presentations might reduce or remove the preferential weighting of the 4th and 8th stimulus items that was seen previously, because observers would no longer be limited by the presence of intrinsic temporal uncertainty. Consistent with this prediction, the pattern of weights across frames in the presence of a gap did not match what was found in the previous experiments. Although different patterns were found for 133 and 400 ms gap durations, neither showed the characteristic preferential weighting for the 4th and 8th stimulus items seen previously, in Fig. 4. However, interpretation of these results is complicated by the fact that the preferential weighting of the 4th and 8th frames was also not obtained with a 0 ms gap – a condition that directly replicates the conditions tested in the previous experiments.

The 0 ms gap condition was identical to conditions in several of our previous experiments. As such, our failure to replicate the results from our previous experiments in the 0 ms condition requires explanation. One possibility is that having been exposed to conditions in which there was a gap between stimulus halves changed observers’ strategies. Recall that the orders in which observers completed the three different gap-durations conditions were counterbalanced, so that only 1/3 of the observers experienced the 0 ms gap width condition before the other two conditions. If exposure to one or both of the non-zero gap conditions prior to the 0 ms gap duration condition was responsible for failure to replicate the results of previous experiments, we would expect observers who were exposed to the 0 ms gap condition prior to any non-zero gap condition to show the same pattern of weights that was seen in previous experiments. To test this possibility, carried out a conditionalized reverse correlation analysis. We first sorted the 24 observers into three categories according to the gap duration with which they were tested first, and then did a separate reverse correlation analysis on each of these three subsets of data. The results of this analysis are shown in the top right and bottom two panels of Fig. 13. Each panel, based on just eight observers, shows the estimated weights for one gap duration condition as a function of an item’s ordinal position within a stimulus. The top right panel plots these data for observers who were initially exposed to a gap duration of 0 ms, the bottom left a gap duration of 133 ms and the bottom right a gap duration of 400 ms. As indicated by the relatively large error bars on each data point, the weights estimated by this analysis are far noisier those generated from the complete data set, because of the reduced numbers of trials comprising each data subset. Nevertheless, it appears that the results of our previous experiments were partially replicated when observers were initially exposed to the 0 ms gap condition and there was either a 0 or 133 ms gap in the stimulus. Under those conditions, the 4th frame of the stimulus sequence received preferential attention, partially mirroring the pattern seen previously.

Although variability of the results from the conditionalized reverse correlation analysis forecloses truly strong conclusions, our results in aggregate are consistent with the hypothesis that temporal uncertainty influences the pattern of weights observers assign to items in a stimulus sequence. Specifically, the conditionalized reverse correlation analysis suggests that experience with a fixed temporal landmark can reduce these effects. This interpretation is consistent with the overall picture painted by the results of the d' analysis – namely, that temporal uncertainty plays an important role in how observers perform when they attempt to discriminate repeated sequences from non-repeated ones.

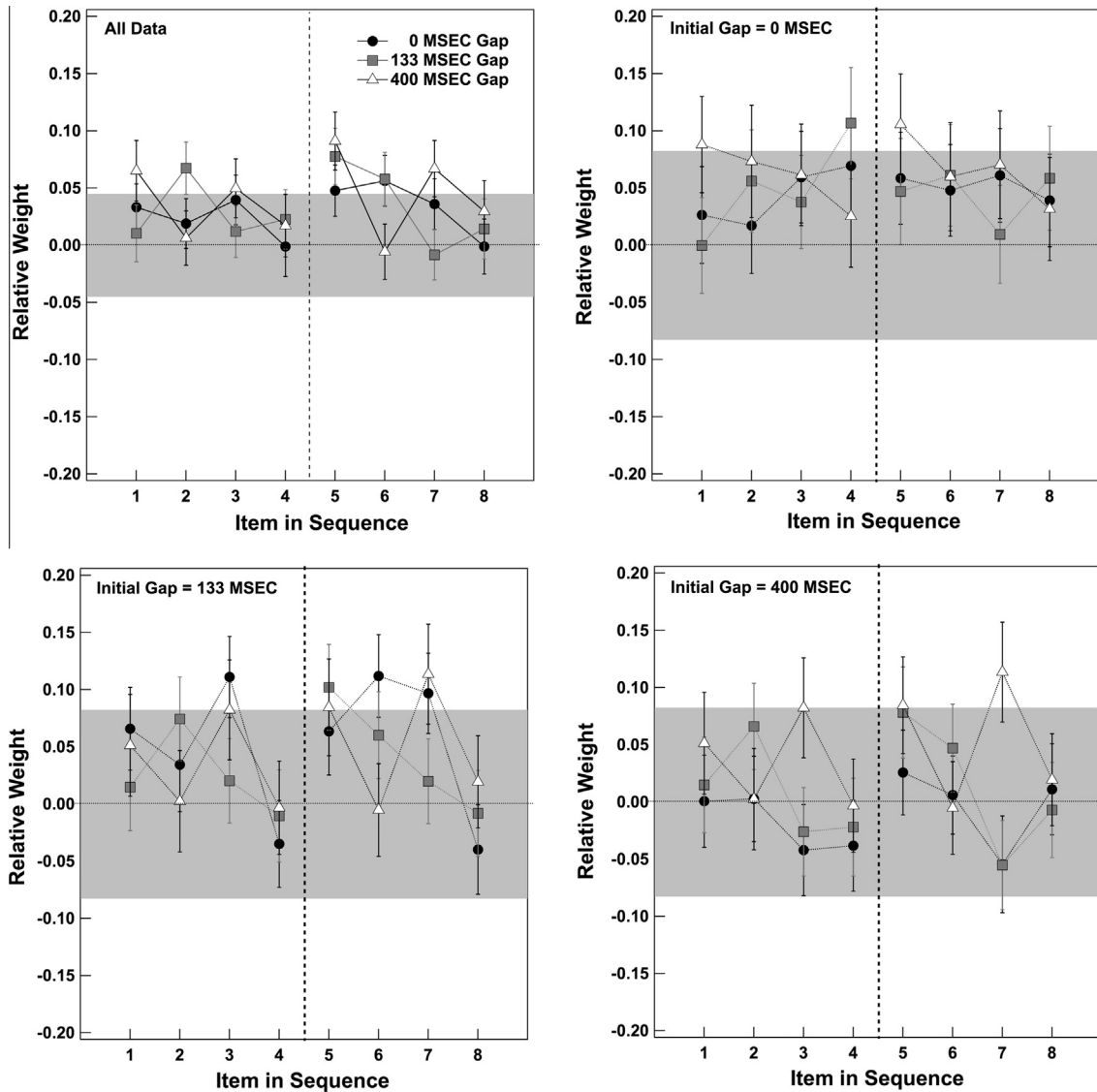


Fig. 13. Mean kernels estimated by reverse correlation for sequences into which various durations of gap have been inserted between items four and five of the sequence. Upper left panel: kernels estimated by averaging over all orders of gaps. Upper right panel: kernels estimated only from observers tested first with 0 ms gap (i.e., no gap). Bottom left panel: kernels estimated from observers tested first with a 133 ms gap. Bottom right panel: kernels estimated from observers tested first with a 400 ms gap. Error bars represent ± 1 s.d., estimated by bootstrap simulations. Gray band in each panel denotes ± 2 s.d. confidence region, estimated by bootstrap simulations.

7. General discussion

By testing observers' ability to discriminate between temporally repeating and temporally uncorrelated random noise sequences our experiments explored some key properties of learning and short-term visual memory. Unbeknownst to observers, our stimuli included some recurring 'fixed' exemplars of repeated noise that appeared multiple times within a block of trials. Across experiments, we manipulated various aspects of the noise sequences, such as their dimensionality, their temporal symmetry, and the duration of a temporal gap inserted at their temporal midpoint. We found that: (1) over the course of relatively few trials, observers were able to learn the properties of both 1D and low-resolution 2D fixed sequences of temporally repeating random visual noise; (2) observers placed more weight on temporal locations that correspond to the mid and end points in a stimulus sequence, which may reflect the influence of temporal uncertainty; (3) overall performance (d') was positively correlated with the total summed contrast of the fixed repeating noise samples; (4) learning of a fixed

repeated noise sample's properties persisted for at least 24 h; and (5) introduction of temporal mirror symmetry severely disrupted observers' ability to discern repeating from non-repeating noise sequences, and prevented learning from taking place. Below, we consider some of the broader implications of these results.

7.1. The correlation between d' and summed contrast

Reverse correlation analyses revealed that every item in a sequence did not exert the same influence on observers' decisions, with the 4th and 8th ordinal positions in a sequence exerting a particularly strong influence. In addition, the examination of total summed contrast within sequences (Fig. 5B) showed that sequences whose value of summed contrast produced the highest d' values. This result is particularly interesting, because there is no necessary relationship between total summed contrast, calculated over an entire eight item sequence, and the recognition of a repetition between the first and second halves of a noise sequence. In fact, the summed contrast of our randomly generated repeated

noise sequences is on average zero, and yet our observers were much better at discriminating repeated noise sequences that were on average more positive than negative in contrast.

Prototype or ensemble-encoding models of cognition suggest one way that summed contrast could influence performance in our task. Models of this class (e.g., Alvarez, 2012) address memory’s well-known capacity limitations by asserting that a set of items or objects is encoded not as individual items, but by means of some summary, scalar value computed over the set of items. It is straightforward to see how ensemble encoding might be relevant in our recognition paradigm. To deal with the high information processing demands of our task, an observer would estimate some ensemble value for the first four items in a sequence, then estimate the comparable value for the last four items, and finally use the relationship between the two estimates as the basis for deciding whether the stimulus sequence had been repeated or not. Given the visual system’s inherent noise and whatever noise might be added in memory, an optimal decision that the sequences were identical should not require a perfect match between the two ensemble codes, that is, zero difference between the two scalars, but only that the values match within some appropriate tolerance limits.

But to which attribute of individual items in our stimulus sequences might ensemble coding be applied? One obvious candidate attribute is summed contrast. After all, we have already seen that a scalar representation of total contrast computed over an entire stimulus sequence affords a reliable separation between FixRN sequences that support good performance and ones that do not (Fig. 5B). We examined whether this summary statistic, computed not over an entire stimulus sequence, but separately for each half sequence, could account for observers’ performance. Because summed contrast values for each half of either a FixRN and RN stimulus sequence always matched one another, those stimuli could not be used to test the idea that ensemble coding might be

at work in our task. But that limitation does not apply to stimulus sequences in the N (non-repeat) condition. Therefore, for every N condition stimulus presented in any experiment, we computed the difference between the summed contrast values in the stimulus’ first and second halves. We then sorted stimuli into quartiles based on those differences in summed contrast values, and computed the mean false alarm rate associated with trials in each quartile.

Fig. 14a displays the proportion of false alarms accounted for by trials in each of four equally populous bins of differences in summed contrast between the first and second half of stimulus sequences. The family of curves represents results from various experiments; the dotted line summarizes results averaged over all the experiments. The leftmost data points represent the one-fourth of trials on which the difference between summed contrasts was smallest; the rightmost data points are for the one-fourth of trials that had the largest difference between summed contrasts. Fig. 14b shows the average results across all experiments, with each quartile’s result expressed as a proportion of all false alarms. These figures suggests there is a substantial systematic relationship between tendency to make a false recognition, on one hand, and the magnitude of the difference in summed contrasts, on the other. For example, Fig. 14b shows that, across all experiments, the trials with the smallest average contrast differences account for 31% of all false alarms, while trials with the largest average contrast differences account for just 17% of false alarms. So, for a non-repeating sequence, the more similar the summed contrast values are for that sequence’s first and second halves, the more likely are observers to misjudge the sequence as having repeated.

As the preceding analysis was preliminary and *post hoc* rather than embedded in a full model, we adopted a default approach for results shown in Fig. 14a and b. Specifically, we ignored the possible influence of noise in encoding and in memory (Sekuler & Kahana, 2007), and the influence of temporal uncertainty, which

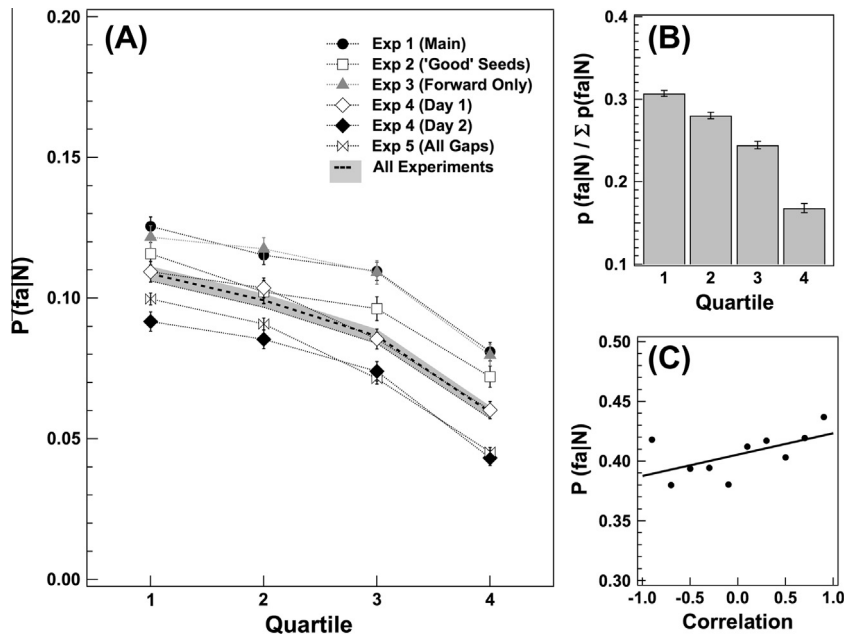


Fig. 14. (A) The relationship between the two halves of a non-repeat (N) stimulus sequence predicts whether the stimulus will attract a false recognition. The probability of a false recognition response to a non-repeat stimulus [$P(\text{fa}|\text{N})$] is plotted against levels of difference between the summed contrast of a stimulus’ two halves. The family of curves represents results from various experiments. Error bars were generated by means of bootstrapping. The dotted line shows results averaged over all the experiments, with size of error bars shown by the gray ribbon. The leftmost data points in each curve represent the one-fourth of trials on which the difference between summed contrasts between sequence halves was smallest; the rightmost data points are for the one-fourth of trials that had the largest difference between first and second halves’ summed contrasts. (B) Results averaged over all experiments, with each quartile’s value expressed as a proportion of all false alarms. Error bars for values shown in the inset were calculated from the fractional uncertainty for each original estimate. (C) $P(\text{fa}|\text{N})$ as a function of the temporal correlation between the first and second halves of the noise sequence (trials were pooled from all experiments). Line through the data is the best (least-squares) linear fit.

was shown in Experiment 5. Moreover, we assumed that all items in a half-sequence entered the computation with equal weight, and that there was a linear relationship between the probability of a false alarm and the numerical difference between a sequence's pair of summed contrast values. Relaxing one or more of these assumptions in future work could well strengthen and clarify the relationship between the variables represented in Fig. 14a and b. In particular, reverse correlation analyses presented earlier showed that the contrast of every individual item probably does not make the same contribution to the aggregate influence of summed contrast. Also, Experiment 3 showed that temporal mirror reversal had an extremely negative impact on performance. Both of these results would only be expected if some aspects of the temporal patterns of the first and second halves of the noise sequences exerted an influence on observers' decisions.

If this second kind of temporal pattern comparison process is in fact taking place, then we would make the strong prediction that observers should be more likely to make false alarms when the temporal patterns of the first and second halves of the pure noise sequences happen to be more similar to each other. We quantified this prediction by first computing the temporal correlation between the contrast values presented during the first and second halves of each N trial noise pattern. This gave us a measure of the temporal similarity between the two halves of each pure noise sequence. Then, each of these correlations was classified as either leading to a correct rejection or a false alarm. Finally, we generated 11 equally spaced temporal correlation bins, spanning between -1 and 1 , and computed the probability of obtaining a false alarm for the trials falling within each bin. Fig. 14c shows the results of this analysis. If observers were basing their decisions in part on the temporal characteristics of the noise within each half of the stimulus, we would expect the probability of a false alarm to systematically increase with greater temporal correlation between the two halves. We did in fact find a moderate but significant linear relationship between $p(\text{fa}|N)$ and temporal correlation (solid line, $r = 0.6$, $p < .01$). Taken together, the results support the idea that observers performed a computation that was a hybrid between a pure ensemble computation, in which multiple items' contrasts are combined into a scalar value, and a computation in which each item in the first half of a sequence is compared against the corresponding item in the second half. Further development and testing of such a hybrid model is beyond the scope of the current report, as it is best accomplished with additional data that are generated with stimulus sequences specially constructed for the purpose (e.g., Rust & Movshon, 2005; Sekuler & Kahana, 2007).

Although observers performed above chance in every one of our experiments, that success may actually represent a lower bound on potential performance in our task. Consider, for example, the algorithm that controlled the quasi-random selection of contrast levels that in our stimulus sequences. As explained earlier, that algorithm was intended to control the contrasts in any one sequence so that individual sequences would not be so distinctive that they would be easily identified, and therefore explicitly recognized if the same sequence appeared on multiple trials. Two components of the stimulus-generating algorithm were particularly important in this regard: the drawing of contrast samples from a low-variance Gaussian distribution centered on the display's mean luminance, and the censoring of extreme values. These constraints meant that a significant proportion ($\sim 10\%$) of all contrast samples lay very near zero contrast, that is, within $\sim 5\%$ of the uniform background luminance upon which contrast sequences were presented. The two constraints also clamped the range of total summed contrast values that comprised any sequence or half-sequence. As both of these variables influence performance (see Figs. 5B and 14), it seems likely that relaxing these constraints would enhance performance.

Finally, there is an additional reason for believing that our results might underestimate the performance that could be achieved, particularly with FixRN stimuli. The intentional homogeneity of our stimulus sequences is likely to encourage retroactive interference (Kahana, 2012), which would retard some of the learning that otherwise could have resulted.

7.2. What is being learned?

In all the experiments reported here, observers had the same basic task: to recognize when a series of quasi-random visual contrasts repeated from the first half of a sequence to the second half. Instructions to observers emphasized matches between corresponding items in a stimulus' halves as the basis on which a sequence should be judged as repeating or not. However, analyses of the connection of performance to stimulus attributes made it clear that, either by choice or as the result of perceptual or memory limitations, observers did not adhere strictly to the item-by-item comparisons implied in the instructions. What, then, did observers learn? More specifically, what were they learning about the FixRN stimuli that allowed the gradual performance improvement we observed?

Our task required observers to pursue two distinct forms of learning. First, observers had to learn to distinguish between stimuli whose halves repeat and stimuli whose halves are uncorrelated. This part of the task would have been made difficult by the influence of noise during encoding and/or in memory. If every item in a sequence had been registered and remembered with perfect fidelity, information embedded in each sequence would have supported perfect discrimination between sequences whose halves repeated and ones that did not. In other words, imperfections in the representation of a sequence, converted what could have been a perfect, deterministic decision into a probabilistic, imperfect one. As a result, over trials observers must develop a criterion for distinguishing between sequences of the two types (Turner, Van Zandt, & Brown, 2011). Then, in parallel with this simple, categorization learning, observers demonstrated that they learned something about the FixRN exemplar in each block of trials. In the remainder of this section, we focus on this second form of learning.

At the outset, one potential explanation of the observed learning can be easily discounted. Experiment Five revealed that temporal uncertainty limited observers' performance. When we reduced observers' uncertainty about the separation between a sequence's first and second halves, performance substantially improved. This result suggests the possibility that over trials, performance with FixRN stimuli might have improved as encoding of timing information became more precise (Buhusi & Meck, 2005; Machado, Malheiro, & Erlhagen, 2009). This increased precision would have diminished temporal uncertainty. As observers showed essentially little or no learning over trials with RN stimuli, it is unlikely that learning with FixRN stimuli arose from some experience-dependent reduction in temporal uncertainty.

In our experiments, improved trial-wise performance with particular FixRN exemplars comprises a form of *incidental* learning. By incidental, we mean learning "which apparently takes place without a specific motive or a specific formal instruction and set to learn the activity or material in question," (McGeogh and Irion, 1952; p. 210). Specifically, observers' task was defined solely in terms of within-trial comparisons between items in the first- and second-half sequences. Moreover, the feedback provided after each response was contingent only whether the observer correctly judged the item-by-item matches within first- and second-half sequences. In other words, observers had "no specific motive or a formal instruction" to treat intermittently-presented FixRN sequences as anything other than one among sequences whose first and second halves matched. Moreover, as the same algorithm

generated RN and FixRN sequences, the statistics of any single stimulus afforded no basis on which observers could correctly classify that stimulus as an RN or FixRN sequence. This could explain why few of our observers recognized that FixRN sequences recurred multiple times – even when directly prompted during debriefing to do so.

As noted above, instructions to observers emphasized the critical role of comparisons between corresponding items in a sequence's halves. However, several results suggest that observers' judgments were based on some subset of all the information represented in a sequence's individual items. For example, reverse correlation analyses (Figs. 4 and 8) revealed that not all items contributed equally. Moreover, the analysis of summed contrast (Fig. 14) demonstrated the importance of aggregate, rather than or in addition to item-by-item representations. We believe that reliance on a limited subset of stimulus information follows naturally from the high-dimensional statistical properties of our stimuli. In our task, an observer is presented with samples from an extremely large number of stimulus states that the observer attempts to map onto simple, binary judgment, a computationally challenging problem often referred to as the "curse of dimensionality". As Botvinick (2012) noted, computational work on learning in high-dimensional state spaces addresses the curse of dimensionality by means of *state abstraction*. In state abstraction, a learner shrinks the space that needs to be explored during learning by collapsing over one or more correlated states. This eases the challenge of learning, though at the cost of information loss. We hypothesize that this beneficial reduction in stimulus information arises early in visual processing, as a result of the rapid rate with which items succeed one another, and the fact that stimulus luminance levels are clustered not far from the mean, background level.

7.3. Relation to other tasks and stimuli

Although the spatial and temporal particulars of the task used in our experiments are distinctive, our observers' success depended upon an ability shared by many studies in psychophysics and memory, namely the ability to recognize that a stimulus has repeated. To take one example, studies of recognition memory ask observers to judge whether some stimulus is "old" (familiar) or "new" (unfamiliar). In studies of visual recognition, observers must judge whether a currently presented stimulus repeats or matches a stimulus or stimuli seen previously.

As mentioned earlier, some aspects of our stimuli and task were modeled after those of Agus, Thorpe, and Pressnitzer's (2010). Psychophysical studies of memory for auditory sequences have a long tradition of using particular repeated sequences as stimuli. The best known of such work was triggered by Julesz's (1962) suggestion that frozen noise stimuli could be used to quantify what he termed a sensory modality's "depth of processing". For Julesz, the term "depth of processing" referred to the longest random signal whose partial periodicities could be detected. For example, one could determine the longest stochastic sequence, either visual or auditory, for which viewers or listeners could distinguish repetitions of a sequence from an entirely random, non-repeating stochastic sequence. When Guttman and Julesz (1963) iterated samples of auditory frozen noise, listeners had no trouble experiencing the distinctive auditory results, so long as samples of frozen noise were no longer than ~1–4 s. In fact, at short repetition periods (e.g., 250 ms), repeated frozen noise signals produced odd, characteristic sounds that listeners likened to the sounds made by a motorboat. Subsequently, others modified this basic paradigm in an attempt to identify the acoustic features that actually supported detection of repetition in frozen noise sequences. For example, Pollack (1972) suspected that listeners exploited a thresholding strategy in order to minimize the amount of information

required by auditory memory. The strategy, he hypothesized, was to detect and compare only the iteration of a sequence's most extreme amplitudes. However, when Pollack greatly reduced that variation among amplitudes present in his sequences, listeners' performance was unchanged. From their ingenious set of experiments, Warren and Bashford (1993) concluded that "...sequences of brief items, whether speech sounds, tones, or stochastic waveforms derived from noise, can be processed globally as complex patterns or temporal compounds, and that resolution into discrete components is not required for discriminating between different arrangements of the acoustic components." The proposition that sequences of auditory frozen noise can be processed globally is at least partially consistent with the finding, described above, that judgments of repetition in visual frozen noise sequences may not depend solely resolving a stimulus sequence into all of its discrete components (see Fig. 14). A detailed examination of this hypothesis will require experiments in which specially constructed stimulus sequences are combined with tasks that rigorously test the degree to which visual frozen noise sequences are processed globally.

It is worth noting that only a few researchers have adapted Guttman and Julesz (1963) basic strategy in order to characterize depth of processing in vision. Moreover, such studies have focused on the detection of repetitions defined over space (Pollack, 1973) rather than on repetition over time, as in our experiments. As such, it is difficult to make principled comparisons between the results of such studies and our own findings.

Finally, we should note that several observers in our experiments volunteered that they had adopted an unusual strategy to help them distinguish repeated from non-repeated sequences of visual contrasts. These observers reported that in their "mind's ear" they translated the sequence of varying contrast into a form of auditory imagery, and then used that imagery as the basis for their judgments (Hubbard, 2010; Seashore, 1938). Of course, no matter how interesting or plausible observers self-reports may be, there is solid evidence that such self-reports may not be valid (Nisbett & Wilson, 1977). However, it may be worthwhile to try to devise some method by which to test observers' claimed reliance on auditory imagery (e.g., Guttman, Gilroy, & Blake, 2005) to solve what on its face is an inherently visual challenge posed by our stimuli and task.

Acknowledgments

We would like to thank two anonymous reviewers for their helpful comments on earlier versions of this manuscript. Supported by CELEST, an NSF Science of Learning Center (SBE-035478), NIMH National Institutes of Health Grant MH-068404, NEI National Institutes of Health Grant EY-019265 and by AFOSR Grant FA9550-10-1-0420.

References

- Agus, T. R., Thorpe, S. J., & Pressnitzer, D. (2010). Rapid formation of robust auditory memories: Insights from noise. *Neuron*, 66(4), 610–618.
- Ahumada, A., Jr., Marken, R., & Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *Journal of the Acoustical Society of America*, 57(2), 385–390.
- Alvarez, G. A. (2012). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131.
- Barlow, H., & Reeves, B. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19(7), 783–793.
- Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Current Opinion in Neurobiology*, 22, 956–962.
- Bruce, V., & Morgan, M. J. (1975). Violations of symmetry and repetition in visual patterns. *Perception*, 4(3), 239–249.
- Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6, 755–765.

- Cattaneo, Z., Mattavelli, G., Papagno, C., Herbert, A., & Silvanto, J. (2011). The role of the human extrastriate visual cortex in mirror symmetry discrimination: A TMS-adaptation study. *Brain and Cognition*, 77, 120–127.
- Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences, USA*, 106(14), 6008–6010.
- Cohn, T. E., & Lasley, D. J. (1986). Visual sensitivity. *Annual Review of Psychology*, 37, 495–521.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science*, 16, 228–235.
- Guttman, N., & Julesz, B. (1963). Lower limits of auditory analysis. *Journal of the Acoustical Society of America*, 35, 610.
- Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, 136, 302–329.
- Jenkins, B. (1982). Redundancy in the perception of bilateral symmetry in dot textures. *Perception & Psychophysics*, 32, 171–177.
- Julesz, B. (1962). Visual pattern discrimination. *Institute of Radio Engineers Transactions on Information Theory*, 8, 84–92.
- Julesz, B., & Hirsch, I. J. (1972). Visual and auditory perception – An essay of comparison. In E. E. David, Jr. & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill.
- Kaernbach, C. (2004). The memory of noise. *Experimental Psychology*, 51(4), 240–248.
- Kahana, M. J. (2012). *Foundations of human memory*. New York: Oxford University Press.
- Kootstra, G., de Boer, B., & Schomaker, L. R. B. (2011). Predicting eye fixations on complex visual stimuli using local symmetry. *Cognitive Computation*, 3(1), 223–240.
- Machado, A., Malheiro, M. T., & Erlhagen, W. (2009). Learning to time: A perspective. *Journal of the Experimental Analysis of Behavior*, 92, 423–458.
- McGeogh, J. A., & Irion, A. L. (1952). *The Psychology of Human Learning*. New York: Longmans, Green.
- Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: weighted sums. *Journal of Vision*, 2(1), 79–104.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Neri, P. (2010). Stochastic characterization of small-scale algorithms for human sensory processing. *Chaos*, 20, 045118.
- Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature of Neuroscience*, 5(8), 812–816.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, 16, 283–290.
- Pollack, I. (1972). Memory for auditory waveform. *Journal of the Acoustical Society of America*, 52, 1209–1215.
- Pollack, I. (1973). Depth of visual information processing. *Acta Psychologica*, 37, 375–392.
- Rust, N. C., & Movshon, A. J. (2005). In praise of artifice. *Nature Neuroscience*, 8, 1647–1650.
- Seashore, C. E. (1938/1967). *The psychology of music*. New York: Dover Press.
- Sekuler, R., & Kahana, M. J. (2007). A stimulus-oriented approach to memory. *Current Directions in Psychological Science*, 16, 305–310.
- Simoncelli, E. P. (2003). Seeing patterns in the noise. *Trends in Cognitive Sciences*, 7, 51–53.
- Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review*, 118, 583–613.
- Visscher, K. M., Kaplan, E., Kahana, M. J., & Sekuler, R. (2007). Auditory short-term memory behaves like visual short-term memory. *PLoS Biology*, 5(3), 662.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, 27, 92–114.
- Wagemans, J. (1995). Detection of visual symmetries. *Spatial Vision*, 9, 9–32.
- Warren, R. M., & Bashford, J. A. Jr., (1993). When acoustic sequences are not perceptual sequences: The global perception of auditory patterns. *Perception & Psychophysics*, 54, 121–126.
- Wilson, H. R. (1980). Spatiotemporal characterization of a transient mechanism in the human visual system. *Vision Research*, 20, 443–452.