Efficient Coding in Visual Short-Term Memory: Evidence for an Information-Limited Capacity

Timothy F. Brady (tfbrady@mit.edu)
Talia Konkle (tkonkle@mit.edu)
George A. Alvarez (alvarez@mit.edu)
Department of Brain and Cognitive Sciences, MIT, Cambridge, MA

Abstract

Previous work on visual short-term memory (VSTM) capacity has typically used patches of color or simple features which are drawn from a uniform distribution, and estimated the capacity of VSTM to be 3-4 items (Luck & Vogel, 1997). Here, we introduce covariance information between colors, and ask if VSTM can take advantage of this redundancy to form a more efficient representation of the displays. We find that observers can successfully remember 5 colors on these displays, significantly higher than the 3 colors remembered when the displays were changed to be uniformly distributed in the final block of the experiment. We suggest that quantifying capacity in terms of number of objects remembered fails to capture factors such as object complexity or statistical redundancy, and that information theoretic measures are better suited to characterizing the capacity of VSTM. We use Huffman coding to model our data, and demonstrate that the data are consistent with a fixed VSTM capacity in bits rather than in terms of number of objects.

Keywords: Visual short-term memory**;** Working memory; Information theory; Memory capacity

Introduction

It is widely accepted that observers are highly sensitive to statistical regularities in the world. This capacity has been used to explain effects from speech segmentation to the emergence of visual objects (Saffran, Aslin & Newport, 1996; Turk-Browne, Isola, Scholl, & Treat, in press). Such regularities also provide an opportunity for memory systems to form more efficient representations by eliminating redundancies. This may be especially important for visual short-term memory, which is known to have a severely limited capacity.

Previous work on VSTM capacity suggests that observers can remember about four objects, independent of the number of features remembered per object (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). In one experiment, observers were shown lines of different colors and orientations. When required to remember either color or orientation alone, they could remember 4 items. Surprisingly, when required to remember both color and orientation, observers could still remember 4 items. In fact, performance was the same when observers had to remember up to four features per object. These data suggested that the amount of information remembered per object is not a limiting factor in memory, and that memory capacity instead depends only on the number of objects to be remembered – consistent with the idea of 'chunks' proposed by Miller (1956) and Cowan (2001).

However, it has become clear recently that there is a serious cost in memory performance for increasing the information content of an object (e.g., objects with multiple colors that need to be stored; Wheeler & Treisman, 2002). This suggests that visual short-term memory (VSTM) cannot hold an unlimited amount of information just because it has been bound to a single object. Alvarez and Cavanagh (2004) proposed an alternate framework that specifically takes into account the amount of information needed to represent each object. They demonstrated that while observers can remember up to four simple objects, they can remember only 1 or 2 complex objects -- presumably because a greater amount of information is required for the complex objects to be remembered well enough to succeed at test. However, because of the nature of the real world objects used in their task, Alvarez and Cavanagh (2004) could not measure the true (information theoretic) information content of their stimuli.

In the present study, we had observers remember color patches because it is possible to exactly quantify the information content of these stimuli in bits (Shannon, 1948). We varied the amount of information per stimulus not by changing the physical appearance of the patches, but by changing the probability of their co-occurrence. Introducing statistical redundancy reduces the amount of information needed to encode the items in the display. This manipulation enabled us to directly compare VSTM models which propose a capacity limit in terms of a fixed number objects versus a fixed amount of information (in bits).

First, we conduct two behavioral experiments, in which we draw stimuli from a uniform distribution (Experiment 1) or from a distribution containing covariance information between presented colors (Experiment 2). Next, using a hierarchical Bayesian model of the learning process and a Huffman encoding scheme, we show that a computational model can predict VSTM performance.

Experiment 1: Uniform Displays

We first assessed the capacity of VSTM for colors drawn from a uniform distribution. This allowed us to get an estimate of the number of bits of color information people can remember under circumstances where no compression is possible.

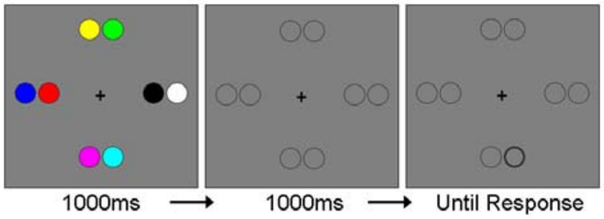

Figure 1: A sample trial. Eight colors are presented and then disappear, and observers have to indicate what color was at a given location.

Method

Eight naïve observers were recruited from the MIT participant pool (age range 18-35) and received 10 dollars for their participation. All gave informed consent.

We presented observers with displays consisting of eight colored circles, arranged in pairs around the fixation point (see sample display in Figure 1). Observers were informed that their task was to remember the locations of each of the eight colors. At the start of a trial, the colors appeared and remained visible for 1000ms. Then the colors disappeared, with placeholder circles present for the next 1000ms (long enough to prevent observers from relying on iconic memory; Sperling, 1960), and then one of the placeholder circles was darkened.

Observers' task was to indicate which of the eight colors had been presented at the indicated location, by pressing one of eight color-coded keys. Observers completed 600 trials, presented in 10 blocks of 60 trials each. Afterward, they completed a questionnaire about the strategies they employed and whether they noticed the presence of patterns in the displays.

The stimuli were presented using MATLAB with the Psychophysics toolbox extensions (Brainard, 1997; Pelli, 1997). The eight colors used were red, green, blue, magenta, cyan, yellow, black and white. The locations of the colors in each trial were chosen randomly, with only the constraint that no color could appear more than once in a given trial.

Results

We estimated the number of objects observers could successfully hold in memory using the following formula for capacity given an eight-alternative forced choice:

$$K = (PC - {}^{1}/_{8}) * (1 + {}^{1}/_{8}) * 8$$

The reasoning behind this formula is that by correcting for chance (1/8) we can examine exactly how many objects from each display observers would have had to remember in order to achieve a given percent correct (PC). Observers in the experiment remembered 3.4 colors on average throughout the experiment (see Figure 2), entirely consistent with previous results on the capacity of VSTM for colors (Luck & Vogel, 1997; Alvarez & Cavanagh, 2004).

Performance on this task can also be quantified in terms of the amount of information remembered. Because there are 8 colors to choose from, each color requires 3 bits of information to encode ($2^3 = 8$). This suggests that the capacity of VSTM for colors drawn from a uniform distribution is approximately 10.1 bits.

These data demonstrate that we can quantify performance both in terms of the number of objects and in terms of the number of bits. In Experiment 2 we introduce covariance information and measure its impact on VSTM. We then model the capacity of VSTM in both uniform displays and in displays with covariance information, to examine whether memory has a fixed information limit in bits.

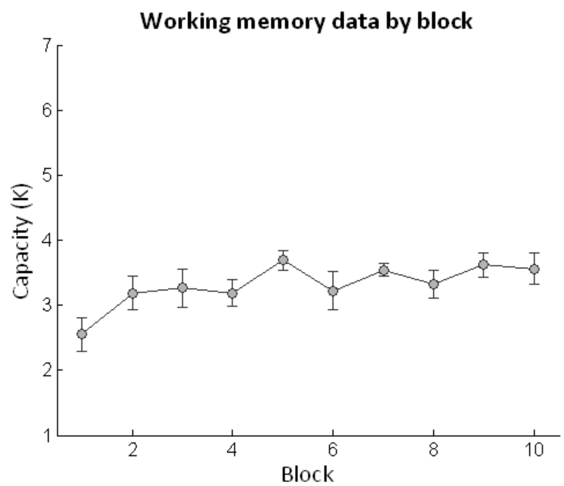

Figure 2: Results of Experiment 1. Error bars correspond to within-subject s.e.m. (Loftus & Masson, 1994).

Experiment 2: Patterned Displays

We next assessed the capacity of VSTM under conditions where statistical redundancy was present in the displays. Specifically, we introduced covariance information between colors. For example, red most often appeared next to green, blue most often appeared next to yellow, etc. If observers are able to learn these pairs, then over time they should be able to remember more items from the display. However, if memory is limited by the number of objects, then performance should be stable over time.

Method

Eight new observers were recruited. Methods were the same as Experiment 1, except that stimuli for each trial were not chosen randomly. First, for each subject a joint probability matrix was constructed to indicate how likely each color was to appear to the left or right of each other color. This matrix was made by choosing four high probability pairs at random (probability = 0.2151), and then assigning the rest of the probability mass uniformly (probability = 0.0027). As in Experiment 1, all eight colors were present in each display. In order to achieve this, the diagonal was set to zero in order to prevent the same color from appearing twice in the same display.

The pairs were constrained so that each color was assigned to exactly one high probability pair. For example, if (Blue, Red) was a high probability pair in this joint probability matrix, the observer would often see blue and red appear together, with blue on the left and red on the right. High probability pairs accounted for approximately 80% of the mass of the probability distribution, and consequently about 80% of the pairs displayed during the experiment.

In the final block of the experiment, the distribution the displays were drawn from was changed to a uniform distribution. This eliminated the regularities in the display, and allowed us to assess whether observers had used the regularities present in the displays to improve their performance.

Results

We found that observers in the patterned condition could successfully remember K = 5.1 colors after learning the regularities in the displays (block 9), significantly higher than the K = 3.1 colors they were able to remember when the displays were changed to be uniformly distributed in block 10 (See Figure 3; $t(7) = 8.30$, $p = 0.00007$).

This suggests that, if we consider working memory capacity in terms of the number of items remembered, observers were able to use the regularities in the displays to increase their capacity past what has been assumed to be a fixed limit of approximately four simple objects. However, one concern is that observers might simply have remembered one color from each pair and then inferred what the other colors were after the display was gone. This would suggest that observers were actually only remembering four objects.

In order to eliminate the possibility that this explicit inference was the only reason for our effect, we separated out trials where the tested item was from a high probability pair from those where the tested item was from a low probability pair. In other words, if blue often appeared with red, we considered only the trials where blue appeared with another color. On these trials, an explicit inference process would cause observers to report the wrong color. When we examine only these trials, we still find that capacity in block 9 is significantly higher than in block 10 (4.9 colors in block 9 and 3.1 colors in block 10; $t(7)=3.63$, $p=0.008$).

We suggest that observers learned to encode the high probability pairs as a single unit using a more efficient representation. This allows them to both hold more high probability items and more low probability items in memory than when the displays were uniformly distributed.

While this is consistent with ideas about chunking (Miller, 1956) as an alternative, we suggest that an information metric might be more useful than the number of items retained. We therefore performed an information theoretic analysis of the data to examine whether observers might have a fixed capacity in terms of bits, rather than number of items. Using the data from both Experiment 1 and 2 allowed us to asses whether, across all blocks of both experiments, observers are remembering the same amount of information, even though they are successfully encoding more colors at some points than others.

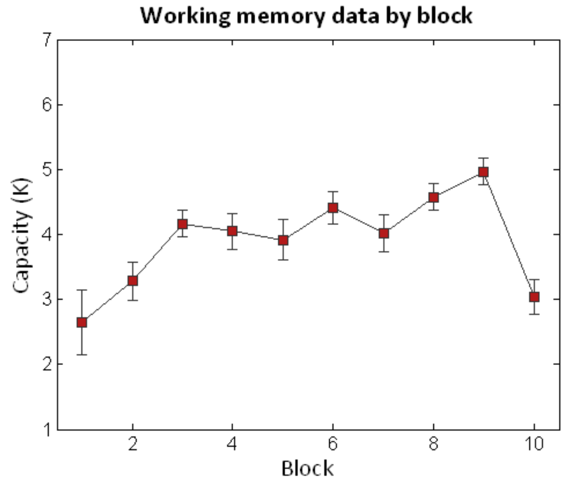

Figure 3: Results of Experiment 2. Block 10 is where the distribution was changed to be uniform. Error bars correspond to within-subject s.e.m.

Modeling

The purpose of the modeling is to test the hypothesis that there is a fixed bit limit on short-term memory. First, we model the learning of the color regularities based on the number of times they saw each pair of colors. Second, we assess how these learned statistics translate into representations in bits, using Huffman coding (Huffman, 1952).

To model the learning of the color pairs, we assumed that observers treated the stimuli as though they were generated from rolls of a 64 sided die, with one side for each possible color pair (Red-Blue, White-Black, Black-White, etc.) We then modeled how they would learn what weight was associated with each side of the die. We did this by counting the frequency with which each pair appeared, plus a prior probability on the die being uniform, using a hierarchical Bayesian model.

Once we had modeled how observers would learn the probability distribution, we next assessed how this probability distribution would help them encode the stimuli more efficiently. We used a compression algorithm, Huffman coding, to see how effectively observers should have been able to represent stimuli that were either drawn from the probability distribution they had learned (blocks 1-9) or drawn from a different distribution (block 10).

Learning the color pairs

We used a Dirichlet-multinomial model to infer the probability distribution that the stimuli were being drawn from, given the color pairs that had been observed. We let d

equal the observations of color pairs. Thus, if the trial represented in Figure 1 is the first trial of the experiment, after this trial d = {Yellow-Green, Black-White, Blue-Red, Magenta-Cyan}. We assume that d is sampled from a multinomial distribution with parameter θ. In other words, we assume that at any point in the experiment, the set of stimuli we have seen so far is a result of repeated rolls of a weighted 64 sided die (one for each cell in the joint probability matrix; i.e., one for each color pair), where the chance of landing on the ith side of the 64 sided die is given by θ_i. Note that this is a simplification, since the experiment included the additional constraint that no color could appear multiple times in the same display. However, this constraint does not have a major effect on the expected distribution of stimuli once a large number of samples has been obtained, and was thus ignored in our formalization.

We set our *a priori* expectations about θ using a Dirichlet distribution with parameter α. The larger α is, the more strongly the model starts off assuming that the true distribution of the stimuli is a uniform distribution (e.g., that the multinomial distribution is using a 'die' that is weighted to land on each possible cell equally). Using statistical notation, the model can be written as:

$$\theta \sim \text{Dirichlet}(\alpha)$$
$$d \sim \text{Multinomial}(\theta)$$

To fit the model to data we set a fixed α and assume that the counts of the pairs that were shown, d, are observed for some time period of the experiment. Our goal is then to compute the posterior distribution $p(\theta \mid d, \alpha)$. The maximum of this posterior distribution is then our best guess at the true probability distribution that the stimuli are being drawn from, and the variance in the posterior indicates how certain we are about our estimate. The posterior of this model reduces to a Dirichlet posterior where the weight for each color pair is equal to the frequency with which that color pair appears in d, plus the prior on that pair, α_i.

This probability model allowed us to infer what observers might believe about the probability of each color pair from the data they had observed. The benefits of a Bayesian model over just counting the frequency with which each pair had appeared are mostly evident early on in the experiment, where observer's prior beliefs come into play most strongly.

Huffman coding

Any finite set of options can be uniquely encoded into a string of bits. For example, if we wished to encode strings consisting of the four letters A, B, C, and D into strings of bits, we could do so by assigning a unique two bit code to each letter and then concatenating the codes. Imagine we had assigned the following codes to the letters: A = 00, B = 01, C = 10, D = 11. The string ACAABAA could then be written as 00100000010000 (14 bits), and uniquely decoded to retrieve the original string.

Importantly, however, this naïve method of generating a code performs quite badly in the case where some letters are much more likely to appear than others. So, for example, if $P(\text{A}) = 0.5$, and $P(\text{B}) = 0.2$, $P(\text{C}) = 0.2$, and $P(\text{D}) = 0.1$, then we can achieve a great deal of compression by representing strings from this language using a different code: A = 0, B = 10, C = 110, D = 111. Using this code, the string from above, ACAABAA, would be represented as 0110001000 (10 bits), a significant savings even for such a short string (29%). Note that it can still be uniquely decoded, because no code is a prefix of any other code.

Huffman coding (Huffman, 1952) is a way of going from the probabilities of a set of symbols to a binary code for representing those symbols in a compressed format (the example codes for A, B, C, D were generated using a Huffman coding algorithm). Here, we used Huffman coding to estimate how much savings observers should show as a result of the fact that the color pairs in our experiment were drawn from a non-uniform distribution.

We used the probabilities of each color pair, as assessed by the hierarchical Bayesian model described above, to generate a unique bit string encoding all of the stimuli from a given block of the experiment. We supposed that if observers were using some form of compression to take advantage of the redundancies in the display, the length of the code that our compression algorithm generates should be inversely proportional to how many objects observers were able to successfully encode. In other words, if there were many low frequency color pairs presented (as in block 10), these items should have longer codes, and observers should be able to successfully remember fewer of them. Alternatively, if there are many high frequency color pairs presented, the better observers' estimate of the true probability distribution, the better they should be able to compress the input.

Results

With these learning and coding models, we can compute a prediction about the memory performance for each subject for each block. In order to assess the fit between the model and the behavioral data, we used the following procedure. For each display in a block, we calculate the number of bits required to encode that display based on the probabilities from the learning model. Next, we correlated the average number of bits per display from the model with the memory performance of the observers. We expect that the *fewer* bits/display needed, the *better* the memory performance, thus we expect a negative correlation.

This prediction holds quite well, with the maximum fit between this Huffman code model and the human data at $\alpha = 100$, where r, the correlation coefficient between the human and model data, is -0.94 ($p = 0.0003$). This large negative correlation means that when the model says there should be long bit strings necessary to encode the stimuli, human VSTM capacity is low, exactly what you would expect if VSTM had a fixed size in bits and took advantage of a compression scheme to eliminate redundant information.

Importantly, this model allows us to examine if there is a fixed-bit limit on memory capacity. The Huffman codes gives a measure of average bits per object, and the memory performance gives a measure in number of objects remembered. Thus, if we multiply the average size of the Huffman code times the number of items remembered, we get an estimate of the number of bits of information a given set of observers recalled in a given block (Figure 4). Notice first that both groups of observers in Experiment 1 and Experiment 2 show the same total bits, despite the overall difference in the number of items remembered between the groups. Second, the total bit estimate remains remarkably constant between block 9 and block 10 in the Experiment 2 group, even though the memory performance measured in number of items showed a significant cost when the statistical regularities were removed. The fact that this estimate is constant across the entire experiment, whereas the estimate in terms of number of objects varies a great deal, suggests that bits are the appropriate way to quantify human VSTM capacity.

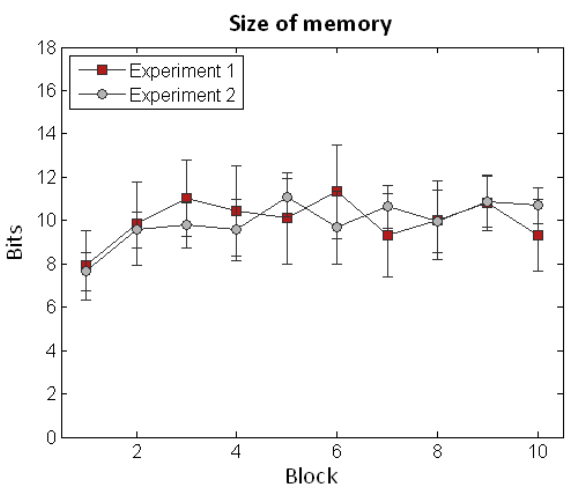

Figure 4: The size of memory estimated in bits, rather than number of objects (using the Huffman coding model). Error bars represent 1 s.e.m.

Importantly, the fit between the human and the model is reasonably good across a broad range of values for the prior, averaging an r value of -0.81 (std: 0.08) in the range $\alpha = 1$ to 200. The fit is poor where the prior is very low, since with no prior there is no learning curve – the model immediately decides that whatever stimuli it has seen are completely representative of the distribution (e.g., like a non-Bayesian model would do). The fit is also poor where the prior is very high, because it never learns anything about the distribution of the stimuli, instead generating codes the entire time as though the distribution was uniform. However, across much of the middle range, the model provides a reasonable approximation to human performance.

Discussion

The current study explored whether the capacity of visual short-term memory is better characterized in terms of the number of objects or the amount of information that can be remembered. We investigated this issue by using information theoretic ideas about compression. Specifically, we introduced covariance information between colors, and asked if VSTM could take advantage of this redundancy to form more efficient representations of the displays. We found that observers could successfully remember 5 colors on the patterned displays, significantly higher than the 3 colors remembered when the displays were drawn from a uniform distribution

These data suggest several conclusions. First, they suggest that VSTM is capable of representing more than four simple objects. In cases where there is statistical redundancy, it is not necessary to encode as much information to represent the objects, and VSTM can represent at least five, and probably more, objects. Together with experiments showing that increasing the amount of information stored per object decreases the total number of objects that can be remembered (Alvarez & Cavanagh, 2004), this suggests that it is the information content, not the number of objects, which is important for understanding VSTM capacity. Measures based on the number of objects fail to capture object complexity, statistical redundancy, and the difficulty of the test comparison, all of which affect the information load for storing objects.

Second, our results suggest that VSTM capacity can be simply modeled by positing a fixed capacity in bits. Across both Experiment 1 and Experiment 2, changes in the estimated capacity of observers are explained by corresponding increases or decreases in the length of the code necessary to represent pairs, resulting in a nearly constant estimate of the total size of memory in bits (about 10, see Figure 4). Thus, under conditions where the compression algorithm suggests people should be able to compress the color pairs into fewer bits, they were able to remember more objects; under conditions where the model achieved little or no compression, they were able to remember fewer objects. Both the Huffman coding algorithm and humans achieved similar levels of compression, with the Huffman code length being 36% shorter in block 9 than 10, and human VSTM capacity being 35% greater in block 9 than 10. This suggests that VSTM is quite good at eliminating the redundant information present in the input.

The fact that an information limit is applicable in VSTM is in line with previous work on the capacity of visual attention. For example, Verghese and Pelli (1992) suggested that attention could be accurately modeled at limited by the information content of stimuli in bits. This combines well with the present results to reinforce the tight coupling between the capacity of attention and the capacity of memory (e.g., Cowan, 2001).

More broadly, our results suggest that people are able to successfully use statistical regularities present in the world to store more objects in VSTM. This is of particular interest given the limited capacity of VSTM. Since the objects and events we use VSTM for in the real world are unlikely to be statistically independent, the number of items we can remember information about may be much larger than is usually assumed. Importantly, however, our results differ drastically from typical results of ‘chunking’ (Miller, 1956) – chunking is usually taken to allow you to store more information in memory ‘for free’, and in fact Miller originally proposed chunking as an alternative to information theoretic models. We instead propose that a more efficient representation of some groups of items may be obtained at the expense of less efficient representation of other, less frequent, items. This suggests that the capacity of memory in bits is constant, and it is how we allocate those bits that changes as a result of learning regularities.

These results also suggest a reason why we might be incredibly sensitive to statistical regularities in the visual world. In particular, a great deal of recent work has focused on statistical learning mechanisms, which are capable of extracting many different regularities with only minutes of exposure and appear to be relatively ubiquitous, occurring in the auditory, tactile and visual domains, and in infants, adults, and monkeys (Brady & Oliva, in press; Conway & Christiansen, 2005; Kirkham, Slemmer & Johnson, 2002; Hauser, Newport & Aslin, 2001; Saffran, et al. 1996; Turk-Browne, Junge & Scholl, 2005). The present results suggests that one of the primary reasons for being sensitive to such regularities might be that it allows us to remember more in working memory by eliminating redundancy in our representations.

Acknowledgments

This research is supported by an NIH/NEI fellowship #F32EY016982 to GAA, and an NDSEG fellowship awarded to T.K.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106-111.

Brady, T. F. & Oliva, A. (in press). Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. *Psychological Science*.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433-436.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 31*(1), 24-39.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-185.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53–B64.

Huffman, D.A. (1952). A Method for Construction of Minimum-Redundancy Codes. *Proceedings of the I.R.E.*, *40*(9), 1098-1101.

Kirkham, N. Z., Slemmer, J. A., Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*, 35-42.

Loftus, G. R. & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, *63,* 81-97.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision,* 10, 437-442.

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

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1-29.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134, 552-564.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (in press). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Verghese, P. & Pelli, D. G. (1992). The information capacity of visual attention. *Vision Research*, 32 (5), 983-995.

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 and Performance*, 27, 92–114.

Wheeler, M.E., & Treisman, A.M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64.