

Colour clustering in visual working memory

Ben Cuthbert (b.cuthbert@queensu.ca)

Centre for Neuroscience Studies, Queen's University
Kingston, Ontario, Canada

Martin Paré (martin.pare@queensu.ca)

Biomedical and Molecular Sciences, Queen's University
Kingston, Ontario, Canada

Dominic Standage (d.standage@bham.ac.uk)

School of Psychology, University of Birmingham
Birmingham, United Kingdom

Gunnar Blohm (gunnar.blohm@queensu.ca)

Centre for Neuroscience Studies, Queen's University
Kingston, Ontario, Canada

Abstract

Visual working memory experiments typically involve asking a subject to memorize several visual stimuli such as coloured shapes, oriented lines, faces, or objects. Computational accounts of recall performance often assume that each stimulus presented in a trial is encoded independently, ignoring higher-level ensemble statistics that have been shown to bias recall and impact task performance. Here, we analyzed data from a delayed estimation task that required the report of all stimuli (6 coloured squares). We found evidence for serial dependencies in within-trial reports, suggesting that participants clustered similarly coloured stimuli together. These dependencies were supported by estimates of the mutual information of within-trial report distributions. We present a non-parametric clustering model to quantify the clustering properties of randomly-generated stimulus arrays. We believe this is a promising data-driven approach to characterizing the statistical properties of experimental stimuli. Together, these results provide further evidence that humans encode ensemble statistics of visual scenes in working memory.

Keywords: visual working memory; mutual information; Dirichlet process; non-parametric clustering; ensemble statistics

Background

The limited capacity of visual working memory (VWM) has been studied extensively using various delayed estimation or match-to-sample paradigms. A great deal of focus has been placed on manipulating the VWM "load" by having participants memorize different numbers of discrete experimental stimuli. There are several standard models of VWM performance that can account for load effects (Ma, Husain, & Bays, 2014), but these models assume that all stimuli are encoded into working memory independently.

The visual world has a rich statistical structure, and recent work has shown that people leverage the statistics of experimental stimuli (broadly referred to as ensemble statistics) to improve VWM performance (Brady & Alvarez, 2011). One way this could be done is by clustering similar stimuli together to reduce redundancy and improve encoding efficiency (Nassar, Helmers, & Frank, 2018).

Evidence for the use of ensemble statistics in VWM has largely come from experiments specifically designed to enable such strategies, but it is possible that people leverage

statistical regularities even when performing more traditional VWM tasks with unstructured stimuli. Here, we re-analyzed publicly available data from a recent experiment where participants memorized and reported 6 coloured stimuli (Adam, Vogel, & Awh, 2017; Figure 1). Despite the fact that colours were randomly generated on each trial, we found evidence that participants grouped similarly coloured stimuli together. We also implemented a non-parametric clustering model to investigate specific ensemble statistics that participants may have used.

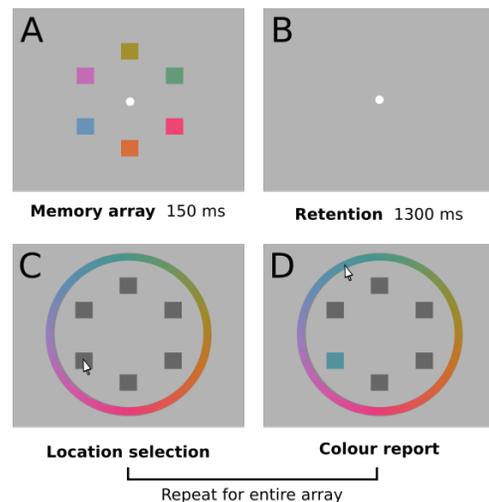


Figure 1: Overview of the whole-report delayed estimation task. (A) Participants viewed 6 coloured stimuli (the memory array) for 150 ms. (B) Blank 1300 ms retention interval. (C) Participants used a mouse to select a stimulus location to report. (D) Participants clicked on the colour wheel to report their memory of the colour at the selected location. C and D were repeated for all 6 stimuli (unspedded), and the order of report was chosen freely by the participant.

Table 1: Notation

Symbol	Meaning
θ	set of true stimulus values
$\hat{\theta}$	set of reported stimulus values
$P(\hat{\theta}_i)$	empirical distribution of the i th report for all trials

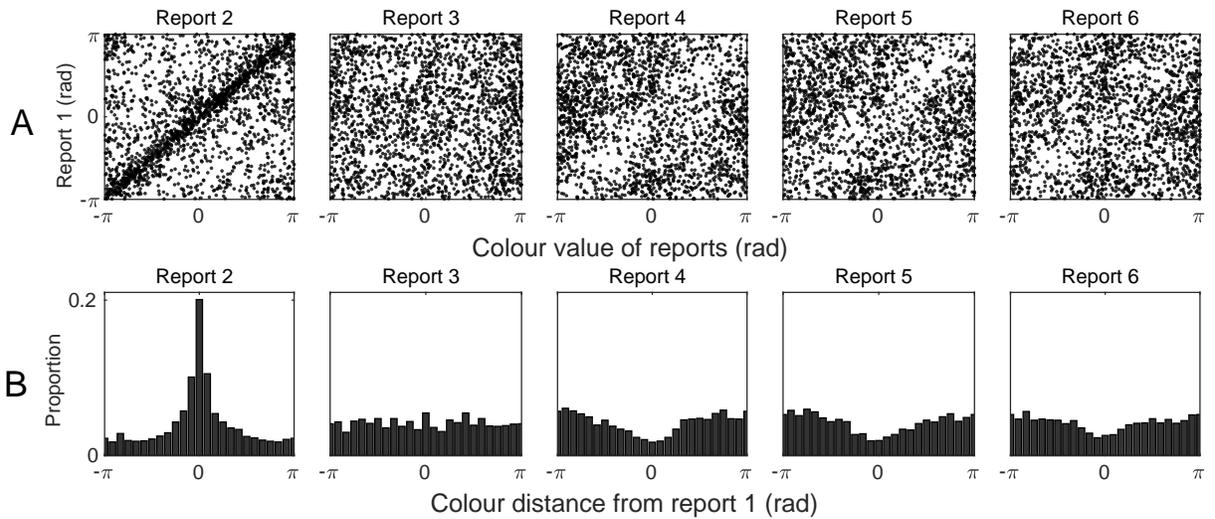


Figure 2: Within-trial joint report distributions. All data are collapsed across participants. (A) Each panel plots the report distribution $P(\hat{\theta}_1)$ against another report distribution $P(\hat{\theta}_i)$. For example, in the top left panel each data point represents the first and second report within the same trial. (B) Each panel plots the distribution of angular distances between the first report and a subsequent report in the same trial. For example, the bottom left panel plots $P(\hat{\theta}_1 - \hat{\theta}_2)$.

Results

Within-trial colour reports are not independent

Adam et al. (2017) analyzed angular report error distributions, and found that angular errors in late reports (ex. the 5th or 6th report at set size 6) were uniformly distributed (Adam et al. 2017). A discrete item capacity limit interpretation of this result is that participants had no information about the final items they reported.

However, report error was only computed relative to the target item being reported, and ignored the trial context (such as other items and reports in the same trial). To test the assumption that item reports are independent, we examined the relationship between reports made within the same trial. On each trial the presented colours θ were uniformly sampled. If all presented colours were encoded independently, within-trial reports should be statistically independent.

Fig 2a includes reports from all trials and participants, where each panel plots a joint distribution of the first report and a subsequent report within the same trial ($P(\hat{\theta}_1, \hat{\theta}_i)$). Similarly, each panel Fig 2b shows the distribution of angular distances between the first report and subsequent within-trial reports ($P(\hat{\theta}_1 - \hat{\theta}_i)$). Immediately consecutive reports (ex 1st and 2nd; Fig 2, column 1) tended to have very similar colour values, while later appeared to be biased away from earlier reports (ex 1st and 5th; Fig 2, column 4). This pattern held for all within-trial joint distributions (not pictured).

These distributions suggest that within-trial reports were not independent. To quantify dependencies between within-trial reports, we used mutual information (MI), a measure of dependence between two random variables that does not assume a particular functional relationship (Cover & Thomas,

2012). For each within-trial joint distribution $P(\hat{\theta}_i, \hat{\theta}_j)$, we computed a mutual information ratio R_1 that estimates the amount of mutual information relative to independent distributions (Fig 3, see Methods). Nearly all joint distributions (and especially $P(\hat{\theta}_1, \hat{\theta}_2)$) contain more mutual information than would be expected if within-trial reports of stimuli in memory were independent.

These results suggest that participants leveraged the ensemble statistics of the arrays to perform the task. In particular, participants appear to have grouped similarly coloured stimuli together.

Non-parametric clustering allows quantification of memory arrays

As discussed above, it has been suggested that information can be pooled across clusters of similar stimuli to improve working memory performance (Brady & Alvarez, 2011; Nasar et al. 2018).

To investigate the impact that colour clustering may have had on task performance, we used a Dirichlet process mixture model (DPMM; Neal, 2000) to characterize the stimulus arrays presented. The DPMM assumes that stimuli on each trial are generated in clusters, and partitions stimulus values θ into K probable clusters based on colour similarity. Critically, the model uses a Dirichlet process as a non-parametric prior on possible clustering structure and therefore avoids making a priori assumptions about the number of clusters present in θ (see Methods).

The DPMM considers every possible partitioning of θ , and provides a posterior distribution over K rather than “hard” assigning the stimuli to specific clusters. Example posteriors for

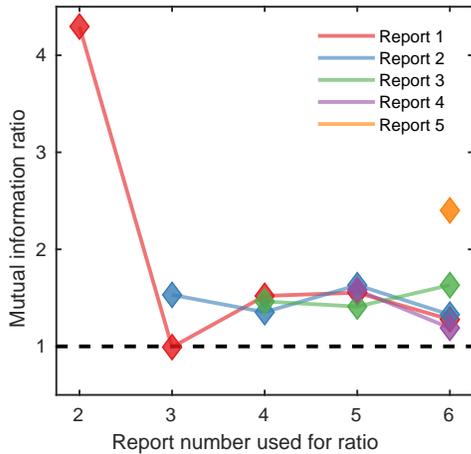


Figure 3: Mutual information ratio of all joint report distributions. The mutual information ratio is an estimate of the mutual information I of $P(\hat{\theta}_i, \hat{\theta}_j)$ relative to the I of two independent distributions (see Methods). Each point corresponds to the R_I for $P(\hat{\theta}_i, \hat{\theta}_j)$ where the colour of the point indicates i and the x-axis indicates j .

three different θ s are shown in Figure 4.

While we are still exploring the parameter space and clustering properties of the DPMM, we believe that this non-parametric approach is a promising analysis method with potentially broad applications. In addition to providing a data-driven quantification of the clustering structure of randomly sampled arrays, DPMMs could also be used to generate stimuli with specific properties. This class of models can also be easily extended to multiple dimensions, and DPMMs have been successfully used for 2-dimensional spatial clustering (Lew & Vul, 2015).

Conclusion

Our results contribute to the growing body of work suggesting that humans leverage the ensemble statistics of visual scenes to aid visual working memory. By considering the joint distributions of within-trial reports rather than individual report errors, we found evidence that people group to-be-remembered stimuli by colour even when the stimuli are randomly generated and presented in a far from naturalistic task setting.

This finding could be strengthened by continued development of the non-parametric clustering model presented here. Data-driven approaches to characterizing visual stimuli have many desirable properties, and in the future could allow for the use of more statistically complex or naturalistic scenes in visual working memory experiments.

Acknowledgments

Financial support for this project was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Foundation for Innovation (CFI).

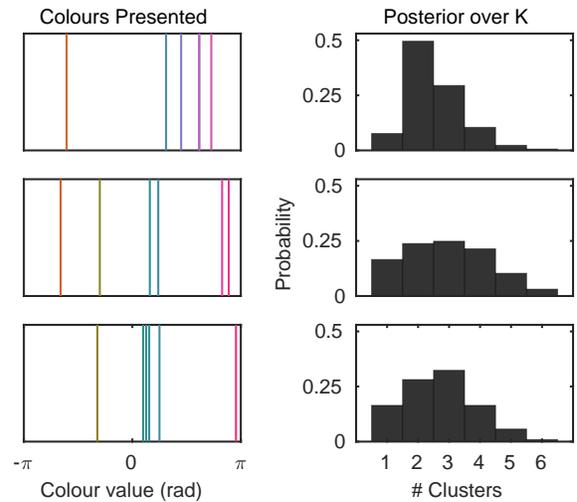


Figure 4: Example trials and DPMM output. Left panels show example colour values θ from 3 trials. Right panels display the accompanying posterior distribution over the number of generating clusters K . In the top trial, the model assigns a high posterior probability that $K = 2$, while in the bottom panel the highest probability is assigned to partitions where $K = 3$. The middle trial includes colours that are more evenly distributed, and as a result the model assigns more evenly distributed probabilities to values of K .

We would like to sincerely thank K Adam, E Vogel, and E Awh for making their experimental data publicly available and therefore for making this project possible.

Methods

Whole-report delayed estimation task

Here, we only considered a subset of the data collected by Adam et al. (2017). Specifically, we restricted analysis to trials with 6 coloured stimuli where participants were permitted to choose their response order. For trial outline and timing, refer to Fig. 1. For complete experimental details refer to the original publication (Adam et al. 2017; Experiment 1a). All data and code for the original publication is available at <http://www.osf.io/kjpnk>.

Colour stimuli The colours presented on each trial were randomly drawn with replacement from a set of 360 colours. The colour set was chosen from equidistant points around a circle in CIEL*a*b* colour space centered at $L = 54$, $a = 18$, and $b = -8$. CIEL*a*b* space was designed for perceptual uniformity, and as such we treat each colour as an angular value along the continuous circular dimension $(-\pi, \pi)$.

Mutual information ratio

The mutual information I of each pair of report distributions $(P(\hat{\theta}_i), P(\hat{\theta}_j))$ was estimated using the standard summation method for two jointly discrete random variables (Cover and

Thomas 2012).

The mutual information ratio R_I was computed by drawing a sample of $U \sim \text{Uniform}(-\pi, \pi)$ equal in size to a given empirical report distribution $P(\hat{\theta}_i)$. The ratio for reports (i, j) is then:

$$R_I(P(\hat{\theta}_i), P(\hat{\theta}_j)) = \frac{I(\hat{\theta}_i, \hat{\theta}_j)}{I(\hat{\theta}_i, U)} \quad (1)$$

where $P(\hat{\theta}_i)$ and $P(\hat{\theta}_j)$ are empirical realizations of the random variables $\hat{\theta}_i$ and $\hat{\theta}_j$. This ratio has an intuitive interpretation: because the uniform random variable U is statistically independent of a given report distribution, the mutual information of $P(\hat{\theta}_i)$ and a sample drawn from U will be very close to 0. If $R_I(P(\hat{\theta}_i), P(\hat{\theta}_j)) \approx 1$, this is evidence that $P(\hat{\theta}_i)$ and $P(\hat{\theta}_j)$ are independent. If $R_I(P(\hat{\theta}_i), P(\hat{\theta}_j)) > 1$, there is evidence for statistical dependency.

In practice, 1000 samples were drawn from U for each comparison to a distribution $P(\hat{\theta}_i)$. The mean of all estimated values $I(\hat{\theta}_i, U)$ was used to compute the ratios in Fig. 3, but using a greater estimate (mean + 1 standard deviation) did not change the reported effect.

Dirichlet process mixture model

Note: Above, $P(\hat{\theta}_i)$ refers to the i th report distribution. This is not to be confused with θ_i , which we will use here to denote a single colour value from the set θ .

A Dirichlet process mixture model (DPMM) was used to infer posterior distributions over all possible clusterings of colours presented in a single trial. The DPMM assumes that the colour values θ in a given trial are sampled from a weighted mixture of infinite components. For a single colour value θ_i the model assumes:

$$\theta_i \sim \text{vonMises}(\mu_i, \kappa_i) \quad (2)$$

where μ_i and κ_i are the mean and precision of the von Mises component that generated θ_i . The probability density function of the von Mises (circular normal) distribution is given by:

$$\text{vonMises}(\theta_i | \mu_i, \kappa_i) = \frac{e^{\kappa_i \cos(\mu_i - \theta_i)}}{2\pi I_0(\kappa_i)} \quad (3)$$

where $I_0(\kappa)$ is the modified Bessel function of order 0.

Rather than estimating a priori the number of components that generated the colours $\theta_{1:6}$ on a given trial, we assume that μ_i and κ_i are drawn from a countably infinite discrete distribution G , which itself is distributed according to a Dirichlet process (DP):

$$\mu_i, \kappa_i \sim G \quad (4)$$

$$G \sim \text{DP}(G_0, \alpha_{\text{DP}}) \quad (5)$$

where G_0 (known as the base distribution) represents the prior over the joint distribution of μ_i and κ_i :

$$G_0(\mu_i, \kappa_i) = \text{Uniform}(\mu_i; -\pi, \pi) \text{Gamma}(\kappa_i; \alpha_\kappa, \beta_\kappa) \quad (6)$$

and α_{DP} is a concentration parameter that influences the distribution of component weights. We initialized $\alpha_{\text{DP}_0} = 1$, which

is the equivalent of a uniform prior over the distribution of weights. For a more detailed treatment of Dirichlet process properties, see Neal (2000).

The DPMM described above is inspired by the DPMM developed by Orhan and Jacobs (2013), but differs in two key ways. First, we have omitted the final level of inference involving noisy observations. Second, Orhan and Jacobs used Gaussian components; here, we have adapted the model and sampling routine for circular von Mises components.

Markov Chain Sampling

Samples from the posterior distribution of the DPMM were generated via Gibbs sampling with auxiliary parameters (Neal 2000; Algorithm 8). This algorithm enables sampling from DPMMs with non-conjugate priors by representing the conditional prior distribution for each observation with ζ auxiliary components. 4000 iterations of the sampler were performed for each stimulus set θ .

Table 2: DPMM Parameters (Figure 4)

Parameter	Description	Value
α_κ	parameter of Gamma prior on κ	2
β_κ	parameter of Gamma prior on κ	1
α_{DP_0}	initial DP concentration parameter	1
ζ	# aux. components (Gibbs sampler)	2

References

- Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, 97, 79–97.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological science*, 22(3), 384–392.
- Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.
- Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the weber noise of relative positions. *Journal of Vision*, 15(4), 10–10.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature neuroscience*, 17(3), 347.
- Nassar, M. R., Helmers, J. C., & Frank, M. J. (2018). Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological review*, 125(4), 486.
- Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2), 249–265.
- Orhan, A. E., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological review*, 120(2), 297.