



Limited memory for ensemble statistics in visual change detection

William J. Harrison^{a,b,c,*}, Jessica M.V. McMaster^a, Paul M. Bays^a

^a Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

^b Queensland Brain Institute, The University of Queensland, QBI Building 79, St Lucia, QLD 4072, Australia

^c School of Psychology, The University of Queensland, McElwain Building 24a, St Lucia, QLD 4072, Australia

ARTICLE INFO

Keywords:

Visual working memory
Short term memory
Ensemble statistics
Change detection
Optimal observer model
Signal detection theory

ABSTRACT

Accounts of working memory based on independent item representations may overlook a possible contribution of ensemble statistics, higher-order regularities of a scene such as the mean or variance of a visual attribute. Here we used change detection tasks to investigate the hypothesis that observers store ensemble statistics in working memory and use them to detect changes in the visual environment. We controlled changes to the ensemble mean or variance between memory and test displays across six experiments. We made specific predictions of observers' sensitivity using an optimal summation model that integrates evidence across separate items but does not detect changes in ensemble statistics. We found strong evidence that observers outperformed this model, but only when task difficulty was high, and only for changes in stimulus variance. Under these conditions, we estimated that the variance of items contributed to change detection sensitivity more strongly than any individual item in this case. In contrast, however, we found strong evidence against the hypothesis that the average feature value is stored in working memory: when the mean of memoranda changed, sensitivity did not differ from the optimal summation model, which was blind to the ensemble mean, in five out of six experiments. Our results reveal that change detection is primarily limited by uncertainty in the memory of individual features, but that memory for the variance of items can facilitate detection under a limited set of conditions that involve relatively high working memory demands.

1. Introduction

Within a second of viewing a scene, humans' ability to remember the precise details of objects is severely limited. Factors that constrain visual working memory have been studied across a large number of paradigms over several decades. While much of this research has focused on the nature of storage for individual items or features (Baddeley, 2012; Ma, Husain, & Bays, 2014) other studies have investigated whether working memory stores visual summary or ensemble statistics in addition to the features themselves (Brady & Alvarez, 2011, 2015a; Brady, Konkle, & Alvarez, 2009; Brady & Tenenbaum, 2013; Corbett, 2017; Nassar, Helmers, & Frank, 2018; Orhan & Jacobs, 2013). Ensemble statistics include, for example, the mean (average) or the variance (spread) of a set of colours. The hypothesis that humans hold such statistics in working memory may suggest the capacity of working memory has been systematically underestimated in many previous studies (Cohen, Denzler, & Kanwisher, 2016).

The notion that ensemble statistics contribute to recall is primarily supported by evidence from experiments in which participants are cued

to reproduce remembered features from a continuum of features (Wilken & Ma, 2004). In one such experiment, Brady and Alvarez (2015a) investigated whether memory performance is influenced by the context in which memoranda appear. They had observers remember one, three or six coloured items, and then reproduce the colour of each item in order of a random spatial cue. Recalled colours were reported by clicking a colour wheel. Importantly, Brady and Alvarez presented the same memory displays to hundreds of participants, allowing them to estimate a distribution of recall errors for each item for each combination of memoranda. They found that, for a given set size, recall errors depended on the context of items, and, within a context, the specific colour value of each memorandum. Some of these effects could be accounted for by a model in which the mean and variance of a set of colours influences the memory of each individual colour in that set.

More recently, Utochkin and Brady (2020) used a similar reproduction task to investigate memory of orientation, while manipulating the range covered by the orientations in each memory array. They found that variability in recall of an individual item was lower when the presented orientations spanned a narrower range, and that estimates were

* Corresponding author at: Queensland Brain Institute, The University of Queensland, QBI Building 79, St Lucia, QLD 4072, Australia.

E-mail address: willjharri@gmail.com (W.J. Harrison).

<https://doi.org/10.1016/j.cognition.2021.104763>

Received 28 October 2020; Received in revised form 2 May 2021; Accepted 3 May 2021

Available online 29 May 2021

0010-0277/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

on average biased towards the mean of the presented orientations. The authors of these studies concluded that storage of ensemble information, specifically the mean and variance of memoranda, facilitates short-term memory for individual items.

While results like these clearly indicate that responses on cued recall tasks reflect more than just memory for the cued item, they do not convincingly demonstrate that ensemble statistics such as mean or variance are stored in memory. This is because the observations in these studies do not discriminate influences of an ensemble statistic encoded into memory during stimulus presentation from combined influences of the individual feature representations, or influences of the same ensemble statistic extracted from the individual memories. For example, a bias in the direction of the mean orientation in a sample array could be due to the summation of individual biases towards each of the non-target orientations in memory, or reflect a bias towards the mean of the individual memory representations. In the first case, the ensemble statistic would not directly contribute to the bias, in the second it would be the source of bias but not have an independent representation in memory. Direct experimental evidence supporting the storage of ensemble statistics is therefore lacking.

In the present study, we used a change detection task to investigate the influence of ensemble statistics on visual working memory by explicitly controlling which statistics could be used to detect a change across stimulus displays. Across multiple experiments we measured change detection sensitivity for colours and orientations, with multiple set sizes and task difficulties. To assess any possible contribution of ensemble statistics, we compared observers' sensitivity with a prediction of the behaviour expected if participants were blind to ensemble statistics and instead optimally combined information across individual items.

Observers viewed two displays of coloured disks (Experiments 1–5) or oriented Gabors (Experiment 6), separated by an inter-stimulus interval of one second. The observers' task was to report whether the items in the second (test) display were the same as, or different from, the first (memory) display. Correct performance on this task required comparing the test display to information held in memory from the first display. On change trials (50% of all trials), test stimuli were generated by shifting the feature value (colour or orientation) of every item in the memory display through a fixed distance in feature space. Importantly, the changes were chosen in such a way as to explicitly control changes in ensemble statistics across displays: in each experiment, we changed the mean of the displayed items, or the variance, or both, and measured observers' sensitivity to detect the changes (d') as a function of these manipulations.

2. Materials and methods

2.1. Participants

120 volunteers completed all trials of at least one of the six experiments (20 participants per experiment; see below). All had normal colour vision and normal or corrected-to-normal visual acuity. The study was approved by the University of Cambridge Psychology Research Ethics Committee and the University of Queensland Medicine, Low & Negligible Risk Ethics Sub-Committee, and all observers gave written informed consent. Nine participants did not complete the experiment for one of the following reasons: they opted to discontinue the experiment halfway through (two participants), they had consistently poor fixation stability (four participants), we were unable to obtain a stable threshold estimate in the pre-experiment procedure after two runs (two participants), or they had a negative d' in the one-item condition (one participant). Three volunteers failed an Ishihara colour vision test and did not participate in the change detection experiments. Data were only analysed for participants who completed all trials. We determined sample size using Bayesian statistics and an optional stopping rule in which we collected data until evidence for main comparisons reached

$BF_{10} > 10$ (strong evidence for a difference) or $BF_{01} > 4$ (moderate evidence for no difference). For each experiment, we first tested 10 participants before analysing data, then tested additional participants until either our stopping criteria were met or we reached 20 participants, for pragmatic reasons.

2.2. General methods

2.2.1. Experimental setup

Participants sat with their head in a forehead and chin rest positioned 57 cm from a calibrated ASUS LCD monitor (1920 × 1200 pixels within an area of 44.8 cm × 28 cm). Fixation position was measured online with an Eyelink 1000 eye tracker (SR Research; 500 Hz). Stimulus presentation was programmed with the Psychophysics Toolbox Version 3 (Brainard, 1997; Pelli, 1997) and Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002) in MATLAB (MathWorks).

2.2.2. Stimuli

Stimuli were coloured disks (4.7° of visual angle in diameter) centred on an imaginary circle (radius 7°) on a uniform grey background (22 cd/m²). The positions of memoranda were randomized from trial to trial with the constraint that neighbouring disk centres were always separated by 90° of arc relative to fixation. The positions of items in the second display were the same as those in the first display. A white spot (0.2° diameter) with a superimposed black crosshair (1-pixel stroke width) was displayed at the centre of the display throughout stimulus presentation. Colours in the memory display were randomly selected from a circle in CIELAB colour space centred at $a = 0$, $b = 0$, with a radius of 40 and fixed luminance $L = 74$ (42.5 cd/m²; calibrated with a Minolta Chroma Meter CS-100). In the following, we express changes of colour in terms of angles of this circle. As in Brady and Alvarez (2015a), the differences between stimuli in a single display were uniformly distributed in the range $\pm 180^\circ$ for coloured stimuli (Experiments 1–5), and $\pm 90^\circ$ for oriented stimuli (Experiment 6).

2.2.3. Procedure

A typical trial sequence is shown in Fig. 1. At the start of each trial an observer had to maintain fixation within 2° of the central fixation point for 500 ms in order for the trial to proceed. If gaze position remained outside of this region for more than 2 s, then a re-calibration procedure was run. After correctly fixating, the memory display was presented for 200 ms (Experiment 1) or 400 ms (Experiments 2–6), followed by a blank interval of one second, and then the test display for the same duration as the memory display. After a 250 ms blank interval, text appeared on the screen instructing observers to press one key if they had detected any colour change, and another key if they had not. If a break in fixation was detected prior to the response screen, a warning message was shown for two seconds, and the trial restarted with the same set of colours (Experiment 1) or new randomly chosen colours (Experiment 2–5) or orientations (Experiment 6).

2.3. Experiment 1

There were 496 trials in Experiment 1, all with two items in the memory display. A single item was presented in the test display in 96 trials, of which 48 were change trials. In the remaining 400 trials, two test items were presented, 200 of which were change trials, including 100 trials in which the mean changed and 100 trials in which the variance changed.

We changed the mean, while keeping variance constant, by rotating both colours through the circular colour space by a fixed angle in the same direction (CW or CCW). We changed the variance, while keeping the mean constant, by rotating the colours in opposite directions. These changes are depicted in Fig. 2. Trials were balanced to include equal proportions of trials in which the variance increased versus decreased. For change trials in which there were multiple items in the study display

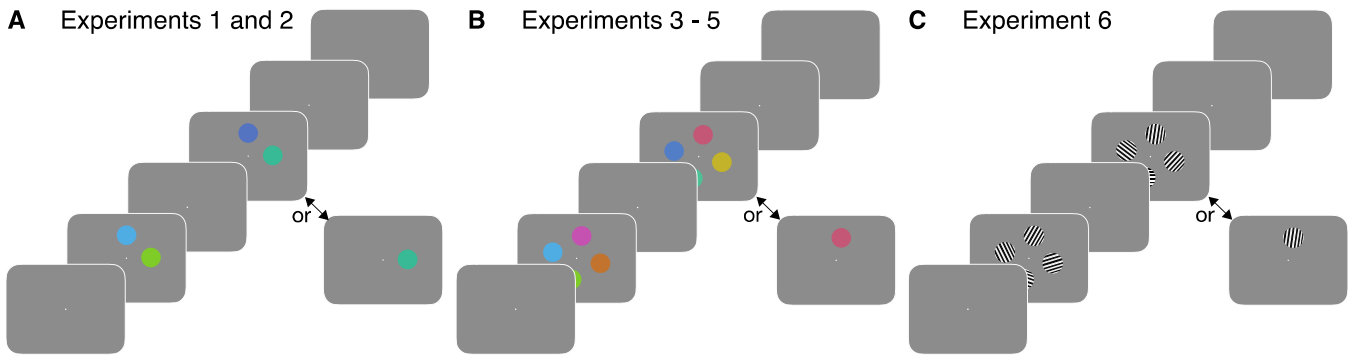


Fig. 1. Design of experiments. Observers reported whether there was a change in two (A) or four (B & C) display items. On change trials, individual colours were chosen to keep the mean the same, while changing the variance (variance-change condition), vice versa (mean-change condition), or both (variance-and-mean-change condition). As shown by the arrows, in each experiment we included a one-item condition to quantify sensitivity to change without the availability of ensemble statistics in the test display. No change occurred in 50% of all trials.

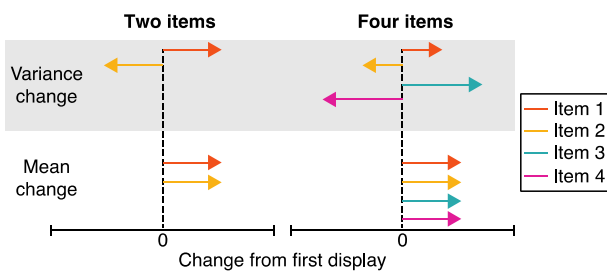


Fig. 2. Schematic of directional changes for each trial type. Observers saw two items in Experiments 1 and 2, and four items in Experiments 3–6. These change directions ensured that the mean was held constant in the variance-change conditions, whereas the variance was held constant in the mean-change conditions. Note that, in Experiment 4, we included a condition in which the variance and mean both changed in place of the variance-change condition shown here (see 2.6 Experiment 4 for details). In Experiment 6, the magnitude of change in all items in the variance-change condition was the same as in the mean-change condition.

but only a single test item was displayed, one disk in the study display was chosen at random and its colour was rotated CW or CCW with equal probability. In all cases, the magnitude of change was fixed, and individually determined for each participant by a thresholding procedure conducted before the main task (see 2.9 *Threshold task* below). Participants completed all trials in a single testing session lasting approximately 90 min. As with all following experiments, trial type (change and no-change) and conditions (variance-change, mean-change and one-item change) were intermixed.

2.3.1. Analyses

We calculated sensitivity to changes as $d' = z(H) - z(F)$ where H and F are the frequency of hits (correct report of a change on a change trial) and false alarms (incorrect report of a change on a no-change trial), respectively. Using signal detection theory, we can make a prediction of the performance in the two-item (mean- or variance-change) conditions we would expect in the absence of ensemble statistics, based on performance in the one-item condition. Assuming optimal summation of evidence from n independent items, the sensitivity to a change is given by:

$$d'_{total} = \sqrt{\sum_{i=1}^n d_i'^2} \quad (1)$$

where $d_i'^2$ is sensitivity to a change in the i th item. Therefore, the predicted sensitivity for mean- and variance-change trials is $\sqrt{2}$ times the sensitivity on single item trials. This formulation is based on the notion

that total evidence of a change is represented by a single point in multi-dimensional space, where each of n -orthogonal axes represents evidence of change in each of n -items. The distance of this point from the origin is d'_{total} , calculated using the Pythagorean Theorem as shown in eq. 1, and as described in detail by Macmillan and Creelman (2004). Note that d'_{total} corresponds to sensitivity of an optimal observer who is blind to ensemble statistics. Observers' empirical d' must therefore exceed this prediction in order to constitute evidence of an influence of ensemble statistics on change detection performance.

Note that eq. 1 is not intended to predict changes in working memory precision under different working memory demands. Instead, working memory load was matched across the one-item and two-item conditions, because the study displays always included two items, and observers had to encode both in order to detect a change in the subsequent display, whether it was a one-item test display or a two-item test display. By computing sensitivity in the one-item condition, therefore, we quantified observers' ability to detect change in a single item, after having encoded two items in memory. Eq. 1 predicts sensitivity to a change in multiple items under the same working memory demands.

All inferential statistics were Bayesian paired samples t -tests calculated using JASP software (JASP Team). We used the default Cauchy prior of 0.707 and all results were robust to standard alternate prior widths. We report Bayes factors in a format depending on which model had the most support, with BF_{01} indicating evidence supporting the null hypothesis (the optimal summation model), and BF_{10} indicating evidence supporting the alternative hypothesis (that observers use ensemble statistics). To assess total evidence for a given hypothesis across multiple experiments, we computed meta-analytic Bayes factors following the methods described by (Rouder & Morey, 2011) using the Bayes Factor package in R (Morey, 2018).

2.4. Experiment 2

Experiment 2 was identical to Experiment 1 except we increased the difficulty of the task by halving the magnitude of change (see 2.9 *Threshold task* below). In Experiment 2, there were 600 trials: 200 trials in which a single test item was presented, 100 of which were change trials; and 400 trials in which two test items were presented (100 change trials in which the mean changed and 100 trials in which the variance changed).

2.5. Experiment 3

Experiment 3 was identical to Experiment 1 except that there were four colours in the memory display which required a different manipulation of change directions (see Fig. 2). On mean-change trials, we rotated all four colours through the same angle in the same direction

(CW or CCW), as in Experiment 1. On variance-change trials, one pair of items (randomly selected) was shifted CW and the other pair shifted CCW. Within each pair the change of one item was twice that of the change of the other item. The result was a change in the variance of the whole set of colours, as well as the variance of every subset of two or more colours, with no change to the mean of the set. To equate difficulty between mean-change and variance-change trials, we used signal detection theory to choose the magnitude of changes. Using Eq. 1, the sensitivity on mean-change trials for 4 items is expected to be twice the sensitivity on one-item trials. To match this sensitivity in the variance-change trials we set the smaller changes to $\sqrt{2/5}$ of the change on one-item trials, and the larger changes to twice this value. Note that, despite using a colour space (CIELAB) intended to be perceptually uniform, our method of equating sensitivities across conditions can only be approximate. We assume, however, that any nonlinearities and individual differences in perceptual sensitivity will average out across trials and participants, and we note that in the majority of experiments, observers' sensitivities are indeed similar across conditions. In Experiment 3, there were 200 trials in which a single test item was presented, 100 of which were change trials (600 trials total for the experiment). Of the remaining 400 trials, four items were presented, 200 of which were change trials.

2.6. Experiment 4

Experiment 4 was identical to Experiment 3, except that we included a condition in which both the mean and the variance changed in place of the variance-change condition. This was achieved by shifting one pair of items (randomly selected) through a large change in the same direction, and the other pair of items through a smaller change in the opposite direction. The change magnitudes were the same as in the variance-change trials in Experiment 3, equating difficulty with mean-change trials.

2.7. Experiment 5

Experiment 5 was identical to Experiment 3 in which we independently manipulated whether the mean or variance changed on change trials, except that we increased the difficulty of the task by halving the magnitude of change (see 2.9 Threshold task below).

2.8. Experiment 6

In Experiment 6, observers' task was to detect a change in orientation of four Gabors. Gabors were sine-wave gratings (1 cycle/degree of visual angle) of maximum contrast in a Gaussian envelope (s.d. = 0.5 degrees). The size of the envelope was chosen to approximate the size of the coloured disks in previous experiments. Stimuli were presented on a 22-in. LED monitor (1080 × 1920 pixels; 60 Hz refresh), with an assumed gamma of 2 (mean luminance = 59 cd/m²). All other details were the same as Experiment 5 with the following exception. In the variance-change condition, the orientations of one pair of Gabors, chosen at random, rotated CW, while the other pair rotated CCW, all with the same magnitude. This produced two minor differences from the previous four-item experiments. First, the magnitudes of change in variance-change and mean-change conditions were equal. Second, while the variance of the whole set changed on variance-change trials, the variance across pairs of items rotating in the same direction did not. In mean-change trials, all four changes were equal, as per the previous experiments.

2.9. Threshold task

Prior to each experiment, each participant completed a change localization task in which we determined a change magnitude that would approximately equate task difficulty across observers (Fig. 3). For Experiments 1–5 we determined observers' colour change threshold, and in Experiment 6 we determined their orientation change threshold. The details of this experiment were identical to those described above except for the following differences. On every trial the number of test items matched the number of memory items, and one item (selected at random) changed colour (Experiments 1–5) or orientation (Experiment 6). There was a 250 ms delay following the offset of the test display, after which white circles outlined the stimulus positions. The observer moved the mouse cursor and clicked on which disk they thought had changed. The magnitude of change was controlled by an adaptive procedure, QUEST (Watson & Pelli, 1983). For the colour experiments (Experiments 1–5), two threshold runs of 40 trials each were interleaved with different starting estimates of $\pi/6$ and $\pi/3$ radians. For the orientation experiment (Experiment 6), starting estimates of the interleaved staircases were 45 and 20 degrees. Data from both staircases were collated and fit with a Weibull function (Experiment 1) or a cumulative Gaussian (Experiment 2–6) to find the change threshold, defined as the midpoint of the function. In Experiments 1, 3, and 4, each observer's threshold was then used

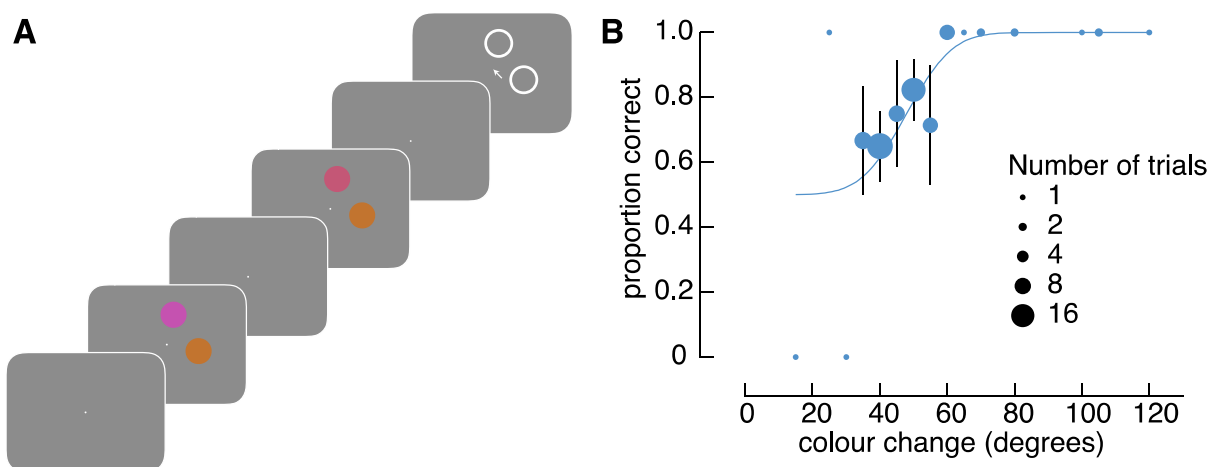


Fig. 3. Threshold task design and example results. A) An example sequence from Experiments 1 and 2. Observers' task was to report which of the two items changed across displays by clicking within the circle surrounding the changed item location. The magnitude of change was adjusted throughout the task according to an adaptive staircase. B) Example data from a single participant in Experiment 1. We fit a psychometric function to the resulting performance data as a function of the magnitude of change. We then used the threshold (midpoint) of this function to set the magnitude of change in the main experiments. In Experiments 1, 3 and 4, change magnitude was set to each observer's threshold, whereas change magnitude in Experiments 2, 5 and 6 was set to half of each observer's threshold.

as the change magnitude in the subsequent change detection experiment. In Experiments 2, 5 and 6, we set the change magnitude to half threshold. Note that the psychometric fits were used only to approximately equate task difficulty in the main experiments across participants, rather than matching theoretical predictions of a particular observer model for this task.

3. Results

In each of six experiments, 20 observers completed a change detection task in which they reported whether or not there was a change in the colour (Experiments 1–5) or orientation (Experiment 6) of items between two displays separated by a one second interval (Fig. 1). There were two (Experiments 1 and 2) or four (Experiments 3–6) memoranda in the first display, and a change occurred on 50% of all trials in the subsequent probe display. As described in the Methods, we generally report Bayes factors in a format depending on which model had the most support, with BF_{01} indicating evidence supporting the predictions of the optimal summation model, and BF_{10} indicating evidence that observers additionally use ensemble statistics to detect changes across displays.

Across all experiments, performance in the “one-item change” condition, in which ensemble statistics could not be compared across displays, was worse than when multiple items were present (all $BF_{10} > 1.5 \times 10^4$, except for Experiment 2 in which $BF_{10} = 5.33$). Such a difference in performance is expected regardless of the use of ensemble statistics because more information is available in multi-item conditions than in the one-item condition (Macmillan & Creelman, 2004).

The critical test of whether ensemble statistics make an independent contribution to change detection is if observers’ sensitivity to a change in multiple items is greater than would be expected were evidence to be summated over individual items. For example, an observer who remembered the mean colour in the memory display in addition to the individual colours would have an advantage in detection on trials where the mean colour was different in the test display. We therefore compared observers’ sensitivity with the prediction of an optimal summation model that does not incorporate information about ensemble statistics. The sensitivity of this model is shown in all results figure panels as a dashed line.

3.1. Experiments 1 and 2: Two-item colour memory

We first tested whether ensemble statistics influence change detection for two coloured items. In Experiment 1, the change magnitude was set to each observer’s pre-determined threshold, whereas in Experiment 2 the magnitude was set to 50% of that threshold (see 2.9 Threshold task). Results are shown in Fig. 4. We found evidence contrary to the notion that observers make use of the mean colour to detect a change in displays: sensitivity in a condition in which the mean changed across displays (orange data points) did not differ from the prediction of the optimal summation model in either experiment (evidence in favour of equal sensitivity, BF_{01} , for Experiment 1 = 3.80; Experiment 2 = 3.96). The optimal summation model, blind to ensemble statistics, therefore accurately predicted performance in the mean-change condition, with a meta-analytic Bayes factor of 8.07, constituting moderate evidence for this model over an observer that used the stimulus mean.

In the variance-change condition of Experiment 1 (blue data point), we found evidence neither for nor against a difference in sensitivity from the optimal summation prediction ($BF_{10} = 1.03$; i.e. the alternative model is only 1.03 times more likely than the null model). In Experiment 2 we found evidence against such a difference ($BF_{01} = 3.08$). Across experiments, the combined evidence was equivocal (meta-analytic Bayes factor = 1.41).

3.2. Experiments 3–5: Four-item colour memory

We investigated whether ensemble statistics are stored in short-term

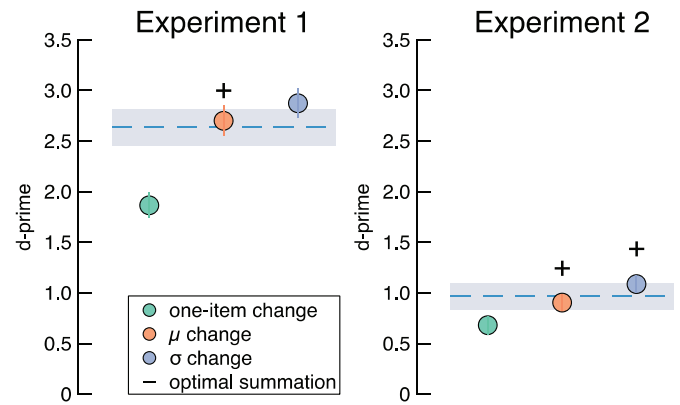


Fig. 4. Results from Experiments 1 and 2. In both experiments, two coloured memoranda were displayed on the first display. In Experiment 1, change magnitude was set to observers’ threshold, and in Experiment 2 change magnitude was set to 50% of threshold (see 2.9 Threshold task). Data points are mean sensitivity for conditions shown in the legend. Error bars indicate ± 1 SE. The ‘+’ symbol indicates a $BF_{01} > 3$ (moderate evidence for observers’ performance matching the optimal summation prediction).

memory under greater memory demands by increasing the number of memoranda to four items. Experiment 3 was in all other ways similar to the preceding experiments. In Experiment 4, in addition to the mean-change condition, we introduced a condition in which we changed the mean and the variance of items in change trials. In both Experiment 3 and 4, change magnitude matched threshold. Experiment 5 included the same mean-change and variance-change conditions as Experiment 3, but change magnitude was 50% of threshold.

We again found evidence contrary to the notion that observers make use of the mean colour to detect a change in displays (Fig. 5). Sensitivity in the mean-change condition across displays (orange data points) did not differ from the prediction of the optimal summation model in any experiment (evidence in favour of equal sensitivity, BF_{01} , for Experiment 3 = 4.2; Experiment 4 = 2.8; Experiment 5 = 3.55). The combined meta-analytic Bayes factor in favour of the optimal summation predictions was 9.8, constituting strong evidence in support of the optimal summation model across experiments.

In Experiments 3 and 4, in which change magnitude matched observers’ threshold, there was evidence for the optimal summation model in the variance-change condition in Experiment 3 ($BF_{01} = 2.9$) and in the mean-and-variance-change condition in Experiment 4 ($BF_{01} = 2.6$). In Experiment 5, in which change magnitude was set to 50% of threshold, however, there was strong evidence that observers outperformed the optimal summation model in the variance-change condition ($BF_{10} = 28.06$).

3.3. Summary of colour experiments

We conducted five experiments in which observers were required to store in memory two (Experiments 1 and 2) or four (Experiments 3–5) items presented in a study display and then report whether those colours changed in a subsequent test display. In change trials, the mean colour, the colour variance, or both, were changed. We found no evidence that observers stored in memory the mean of colours, regardless of the number of memoranda or the colour change magnitude. Indeed, in all experiments we found evidence that detection performance in the mean-change condition was predicted by an optimal summation model blind to ensemble statistics. Importantly, we found strong evidence that observers’ sensitivity exceeded the optimal summation prediction in the variance-change condition of Experiment 5. This result reveals observers did indeed store the variance of colours in working memory and used it to detect change, but not when there were only two memoranda (Experiments 1 and 2), nor when the change magnitude was relatively high

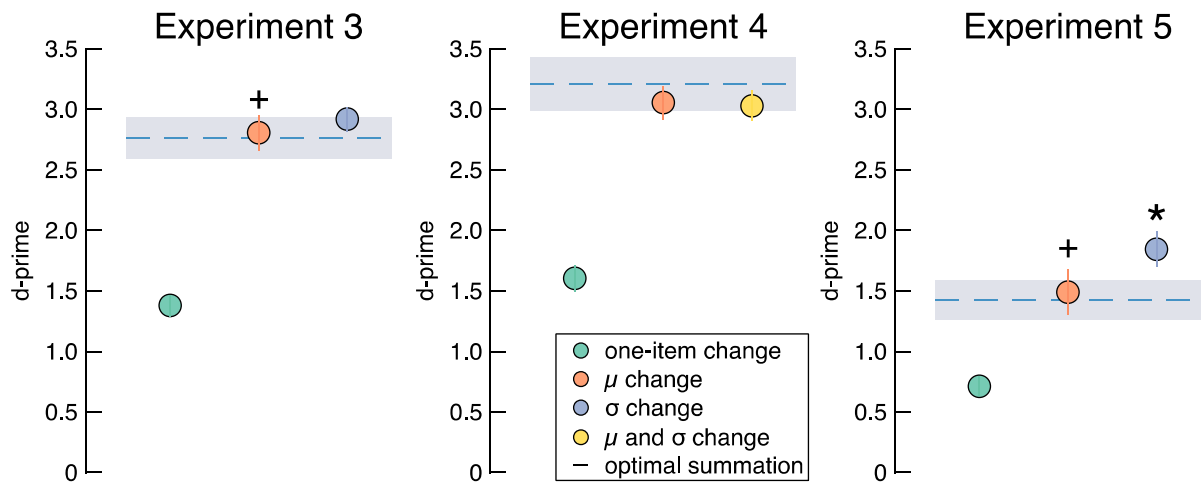


Fig. 5. Results from Experiment 3–5. In all experiments, four coloured memoranda were displayed on the first display. In Experiments 3 and 4, change magnitude was observers' threshold, and in Experiment 5 change magnitude was 50% of threshold. Data points are mean sensitivity for conditions shown in the legend. Error bars indicate ± 1 SE. The '+' symbol indicates a $BF_{10} > 3$ (moderate evidence for observers' performance matching the optimal summation prediction). The '*' symbol indicates a $BF_{10} > 10$ (strong evidence that observers outperformed the optimal summation model).

(Experiments 3 and 4).

3.4. Experiment 6: Four-item orientation memory

In Experiment 6, we investigated short-term memory of oriented items, and tested whether the mean orientation or the orientation variance contributes to change-detection performance. As in Experiment 5, in which we found strong evidence of an influence of the colour variance, we set the magnitude of change to half of each observer's threshold. We again found strong evidence that observers' detection performance in the variance-change condition was greater than that predicted by the optimal summation model ($BF_{10} = 10.7$; Fig. 6). The result of the variance-change condition of Experiment 6 thus replicates that of Experiment 5, but for oriented memoranda instead of coloured memoranda. Unlike with coloured memoranda, however, we also found weak evidence that detection performance in the mean-change condition exceeded the optimal summation model ($BF_{10} = 2.8$).

3.5. Quantifying the contribution of ensemble statistics to change detection sensitivity

In Experiments 5 and 6 we found strong evidence that observers use the variance to detect changes across displays. We can estimate sensitivity to a change in the variance of items using the same evidence summation principle as Eq. 1:

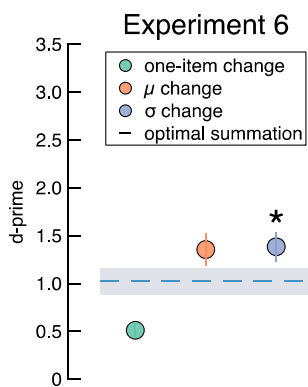


Fig. 6. Results from Experiment 6. Four oriented Gabors were shown on the first display, and change magnitude was set to half threshold. Symbols are as described in Fig. 5.

$$d'_{total} = \sqrt{\sum_{i=1}^n d_i'^2 + d_{ens}'^2} \quad (2)$$

where d'_{ens} is observers' sensitivity to a change in the ensemble statistic. Because we determined d'_i from the one-item condition and d'_{total} from the multi-item conditions, we can calculate d'_{ens} by re-arranging eq. 2 as follows:

$$d'_{ens} = \sqrt{d'_{total}{}^2 - \sum_{i=1}^n d_i'^2} \quad (3)$$

Finally, we calculate w_{ens} , the proportional weighting of the ensemble statistic relative to the other items:

$$w_{ens} = d_{ens}'^2 / d_{total}'^2 \quad (4)$$

Note that the denominator normalises the weights such they sum to one, and, by using the square of the sensitivities, we are weighting by inverse variance according to optimal observer principles. In Experiment 5, the proportional weight given to memory for colour variance versus memory for all the individual colours combined is 0.4: 0.6, and in Experiment 6 the proportional weight for orientation variance versus individual orientations is 0.45: 0.55. These results reveal that observers in these two experiments gave almost the same weight to changes in the ensemble statistic as they did to changes in all other items combined.

3.6. Exploratory analyses

We next addressed the possibility that ensemble statistics are used differently under different display configurations, such as when memoranda are similar or different in colour. We tested change sensitivity when memoranda are high or low in variability (e.g. Utochkin & Brady, 2020) by splitting data from each condition of each experiment. For Experiments 1 and 2, in which there were two memoranda, we separated trials according to whether the difference between colours in the memory display was less than or greater than $\pi/2$ radians (90°). For Experiments 3–6, in which there were four memoranda, each observer's trials were separated according to a median split, i.e. each trial was sorted according to whether the circular standard deviation of items on that trial was lower or higher than the median circular standard deviation across all trials.

The results of these exploratory analyses are shown in Table 1, and in Fig. 7 alongside the main analyses. In general, they are highly consistent

Table 1

Bayes factors for exploratory analyses, expressed as BF_{01} (evidence against observers outperforming the optimal summation prediction).

		Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6
Low variability	Mean-change	4.26	3.72	4.17	2.30	4.13	0.37
	Variance-change	2.97	2.70	4.28	1.50	0.15	0.12
High variability	Mean-change	2.78	4.28	3.80	4.00	3.35	1.48
	Variance-change	1.39	3.93	1.01	4.24	0.44	0.75

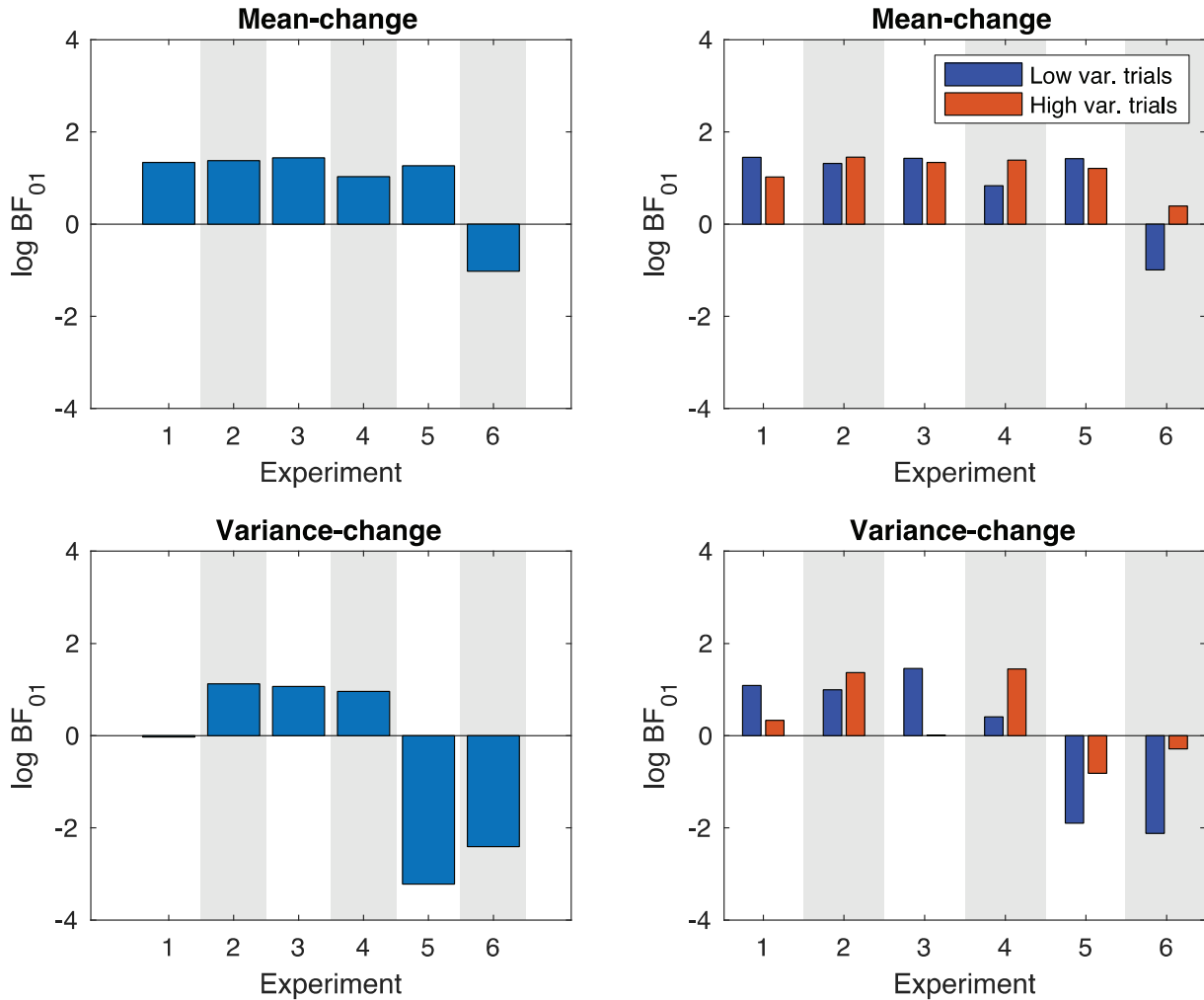


Fig. 7. Summary of all results. Log Bayes factors are expressed such that positive values indicate evidence for the optimal summation model blind to ensemble statistics, and negative values indicate evidence for sensitivity to ensemble statistics.

with those reported above. Sensitivity in the mean-change condition across colour experiments (Experiments 1–5) was again well predicted by the optimal summation model, with a meta-analytic Bayes factor of 12.59 (minimum 2.3) for low variability colours, and a meta-analytic Bayes factor of 8.93 (minimum 2.78) for high variability colours. In the variance-change condition of Experiments 1–4, evidence favoured the optimal summation model regardless of colour variability, but the range of BF_{01} was greater than that for the results from the mean-change condition, ranging from weak anecdotal (minimum $BF_{01} = 1.01$) to moderate (maximum $BF_{01} = 4.28$). That is, regardless of whether there was high or low variability in the colours, we found no evidence in support of the involvement of ensemble statistics in change detection in Experiments 1–4.

The results of this exploratory analysis for Experiments 5 and 6 are again similar to the results reported above. In the variance-change condition when variability between items was low, we again found that evidence favoured a difference from the optimal summation model

in Experiments 5 and 6 ($BF_{10} = 6.67$ and 8.33 , respectively). When variability between items was high, however, evidence for this difference was weak ($BF_{10} = 2.27$ and 1.33 , respectively). We therefore conducted another set of exploratory analyses in which we compared sensitivity across low and high variability trials within each condition of each experiment to assess whether sensitivity was greater in low variability trials compared with high variability trials. Contrary to this possibility, however, observers’ sensitivity was closely matched across low and high variability trial types in all conditions of both experiments, with all Bayes factors favouring no difference (all $BF_{01} > 3$).

In Experiment 5, evidence favoured the optimal summation model in the mean-change condition, regardless of whether variability was low ($BF_{01} = 4.13$) or high ($BF_{01} = 3.35$). In Experiment 6, in which memoranda were oriented Gabors, we found weak evidence for observers outperforming the model predictions when variability was low ($BF_{10} = 2.7$), but not when variability was high ($BF_{01} = 1.48$).

We summarise the results of all experiments and subsequent analyses

in Fig. 7. The results of these extended analyses provide additional confirmation that ensemble statistics contributed to observers' sensitivity to detect change in only a limited number of conditions as described above. These analyses further suggest that sensitivity to changes in colour or orientation variance was largely confined to trials on which there was relatively low variance among items in the initial memory display. Similarly, the weak evidence for sensitivity to a change in mean orientation in Experiment 6 appears to be restricted to trials with low orientation variance. These results are consistent with previous observations that the ability to extract ensemble statistics from a set of stimuli depends on their physical range (Dakin, 2001; Im & Halberda, 2013; Maule & Franklin, 2015; Solomon, 2010; Utochkin & Tiurina, 2014). Although observers may have been motivated to encode ensemble statistics were they given explicit task instructions to do so, our data reveal that such encoding is neither automatic nor obligatory.

4. Discussion

In six change detection experiments we manipulated which ensemble statistics changed across displays to test explicitly the hypothesis that ensemble statistics are stored in short-term memory. We created a prediction of observers' sensitivity under the assumption that their reports were independent of ensemble statistics by deriving an optimal observer model that summates evidence across independently stored memoranda. Evidence that ensemble statistics are stored in memory is therefore any case in which observers out-perform the optimal summation model. We found strong evidence that observers can indeed store and use information about the variance or range of stimuli in a display, but only under the very specific circumstances of detecting small changes to larger stimulus displays where the range of stimulus values is relatively narrow. Furthermore, our results contradict the hypothesis that memory for the mean colour in a display contributes to change detection sensitivity (with weak and limited support for memory of orientation means). Our results thus challenge the suggestion that the ensemble mean plays an important and automatic role in the storage and recall of a visual scene (Brady & Tenenbaum, 2013; Cohen et al., 2016).

We tested change detection performance using two or four memoranda, well within the range of items previously suggested to be influenced by ensemble statistics (Brady & Alvarez, 2015a; Utochkin & Brady, 2020). For detection of a change in two coloured memoranda (Experiments 1 and 2), we found positive evidence *against* the notion that ensemble statistics are stored in memory in three out of four conditions, and equivocal evidence in the fourth condition. These findings suggest that the mean colour or colour variance of two items is not stored in short-term memory, in contrast to the conclusions of some previous studies. We outline possible explanations for this discrepancy below. Importantly, the results from the two-item experiments argue against the notion that ensemble statistics are automatically stored in memory, and this is the case regardless of whether observers are detecting a relatively small or large change of colours, or whether there is low or high variability among memoranda. We cannot rule out the possibility that observers encoded the ensemble statistics in all conditions but did not compare them to the test displays, but we think this is unlikely. Such an account would be hard to reconcile with evidence that observers used the memory for variance in Experiments 5 and 6, but not in any other experiment.

Across four experiments in which observers' task was to detect a change in four items, we found other factors that constrain which ensemble statistics are stored in memory. When the magnitude of change was set to the threshold for localizing a change to a single item (Experiments 3 and 4), we found evidence against the mean colour, colour variance, or both, influencing detection performance. This result reveals that, even with greater short-term memory demands, ensemble statistics are not used in change detection tasks automatically. Only when the magnitude of change was set to half the localization threshold did we find an influence of ensemble statistics (Experiments 5 and 6). In

these experiments we found strong evidence that observers' performance was greater than the optimal summation model in the variance-change condition, and this was true for both coloured disks (Experiment 5) and oriented Gabors (Experiment 6). Higher-order stimulus statistics can therefore be stored in memory, though the lack of an effect of the items' variance in any of the preceding experiments reveals that storage of item variability is not a fundamental principle of short-term memory.

We estimated that, in the conditions where it was used, sensitivity to variance of colours and orientations contributed 40–45% of the evidence used to detect changes in 4-item displays. These estimates suggest that the variance of features contributed to change detection more than any individual item, and almost as much as all items combined. As suggested by Brady and Alvarez (2015b), the contribution of such ensemble statistics to change detection sensitivity could have serious implications for historical estimates of working memory capacity derived from similar tasks. However, this issue is complicated by the fact that ensemble statistics are used in only some conditions, and so a simple rule cannot be applied to correct estimated individual-item sensitivity across different display arrangements. Future studies aiming to investigate solely the storage of individual features can avoid this potential contamination by memory for the variance by using the mean-change condition in our study, which we found does not influence change detection. We can only confidently recommend this control for coloured memoranda, because we found weak evidence that the mean may contribute to change detection sensitivity with oriented memoranda when orientation variance is low.

It remains unclear why observers in our Experiments 3 and 4 did not benefit from a change in the colour variance of four items. These experiments were almost identical to Experiment 5, in which we found strong evidence that the variance of four coloured items contributed to change-detection performance. The only difference between these experiments was that, in Experiments 3 and 4, the magnitude of change was set at observers' localization threshold, whereas in Experiment 5 change magnitude was half threshold. One possibility is that sensitivity of participants in Experiments 3 and 4 was sufficiently high that they were not motivated to store or use the variance of items to detect changes. We performed simulations to confirm it was possible for participants to have outperformed the optimal summation predictions in all experiments, and found this was indeed the case. We simulated the lowest error rates for which a d' -prime could be calculated, one miss and one false alarm per participant, and compared the resulting sensitivity ($d' = 4.65$) to the optimal summation predictions for each experiment using one-sample Bayesian t -tests. We found that observers could have outperformed the predictions in all experiments (all $BF_{10} > 7 \times 10^3$). Indeed, even with three times as many errors (resulting in $d' = 3.78$), observers would have convincingly outperformed the optimal summation model (all $BF_{10} > 3$). These simulations indicate that participants could have used a memory of the ensemble statistic to improve their performance even in the experiments with larger change magnitude, supporting our conclusion that extraction and storage of ensemble statistics is not automatic. Note also that a ceiling effect could not explain why observers did not make use of the colour variance in Experiment 2, in which participants had the lowest levels of sensitivity of any experiment.

Contrary to the suggestion that the mean of a set of colours is automatically stored in memory, we found evidence against storage of the mean in all five colour experiments, with a combined BF_{01} of 637 across experiments. It is important to note that this is not a null effect. This finding instead reveals that, when the variance of colours is held constant, observers' detection performance is best described as the optimal summation of evidence over individual items. This finding cannot be accounted for by a ceiling effect limiting observers' ability to exceed the optimal summation prediction, for the reasons above, and also because observers were able to out-perform the optimal summation model in the variance-change condition in Experiments 5. We found some weak evidence that the mean of four oriented Gabors contributed to

performance in Experiment 6, raising the possibility that different ensemble statistics are encoded for different feature dimensions. Indeed, there is a large body of evidence demonstrating that the mean orientation can be extracted from a set of Gabors (e.g. Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Solomon, 2010). In line with our Experiment 6 results, Solomon found that orientation variance can be detected with greater efficiency than changes in the mean orientation. However, when our analyses were restricted to trials in which there was high variability between oriented memoranda, we found evidence against storage of the mean. Taken together, these findings are inconsistent with automatic storage of the average feature value in a display.

Our study displays were designed to be similar to a previous report that suggested that the mean and variance of colours influence working memory reports (Brady & Alvarez, 2015a). Using a method of adjustment, Brady and Alvarez not only found substantial variability in recall precision for individual items within and across displays, but also that a model incorporating ensemble statistics could account for some of this variability. The principle of this hierarchical model (a variation on one proposed by Orhan & Jacobs, 2013) is that recall estimates of individual features are calculated on the basis of a generative model in which the display may contain clusters of items that are all drawn from a normal probability distribution with a particular mean and variance. This assumption (which does not correspond to how displays are actually generated in these experiments) leads to biases in estimates of a cued item that depend on other items in the display. However, the observed biases taken as evidence for this model do not demonstrate independent storage of ensemble statistics, because these statistics could be estimated at any time from the values of the component items (either visible or in memory). Indeed, in the implementations of the hierarchical model as reported by Brady and Alvarez (2015a) and Orhan and Jacobs (2013), there is no explicit memory for ensemble statistics, and inference is based only on independent noise-corrupted memories of each of the individual features in the display. In contrast, to see an advantage over the optimal summation model in the present study, it would be necessary for participants to explicitly store a summary statistic independent of the individual item colours or orientations.

Our optimal summation predictions were based on extrapolating from performance in conditions where only one of the items in the memory array was presented again in the test display. On these trials, the ensemble statistics that observers might have stored from the memory array were not present in the test display for comparison, which justified our using performance on these trials as a baseline measure for an observer model that is blind to ensemble statistics. However, we do not rule out the possibility that memory for ensemble statistics could have contributed indirectly to performance on these trials. Specifically, observers could in theory evaluate the single item in the test display as if it were a random sample drawn from a probability distribution parameterized with the ensemble statistics remembered from the sample array. If the single feature available at test had low probability under this distribution, it would provide some evidence for a change having occurred that could supplement evidence based on the individual memory for that specific item. As a result, our calculated optimal summation prediction for the mean- and variance-change trials would to some degree overestimate the performance of an observer with no access at all to ensemble statistics from the memory array. Critically, however, it would still predict poorer change detection performance compared to an observer that could directly compare the mean and/or variance of features in the memory and test displays, and would therefore still provide a valid baseline against which to test if our participants have access to those ensemble statistics.

Our optimal summation model formalises a strategy in which observers encode all items in the study display independently, compare their memory of each with the spatially corresponding item in the subsequent test display, and then base their response on the summation of evidence (Macmillan & Creelman, 2004). Although, in principle, observers could have adopted a strategy of storing only a subset of the

displayed items and still detected changes in the mean- and variance-change conditions, we think this is unlikely for two reasons. First, summation of information across multiple items with independent noise is the optimal decision rule providing the greatest sensitivity to detect changes on these trials. Second, this strategy would be highly sub-optimal with respect to the single-item trials, which were unpredictably interleaved with the other conditions, because on the proportion of trials in which the probed item was not stored the observer would have no information on which to base their response.

While we found conditions that greatly constrain the storage of ensemble statistics, we think it is plausible that such statistics may be more readily stored under circumstances that strongly promote the spatial grouping of memoranda (Brady & Tenenbaum, 2013; Victor & Conte, 2004). However, even when we restricted our analyses to include only trials in which memoranda were similarly coloured, thus promoting grouping by similarity (Wagemans et al., 2012), we again found no effect of the mean or variance in Experiments 1–4. Tasks might also be designed to encourage storage of ensemble statistics by making them more difficult to solve using individual item information, for example by unpredictably changing item locations between memory and test displays.

We do expect ensemble statistics to be stored in situations in which individual features cannot be perceptually distinguished, as in visual crowding, where observers may perceive only the average of a set of features belonging to closely-spaced objects in the peripheral visual field (Balas, Nakano, & Rosenholtz, 2009; Dakin, Cass, Greenwood, & Bex, 2010; Harrison & Bex, 2015; Parkes et al., 2001). Such crowding could occur in memory experiments with sufficiently large set sizes, and under these conditions, ensembles may indeed be stored in memory instead of individual items. However, this would reflect a perceptual limitation rather than a feature of working memory.

5. Conclusions

When trying to detect a change in a visual scene, our experiments strongly suggest that people do not always compute or remember the mean or variance of items within the scene. However, memory for the variance of colours or orientations can facilitate change detection under a limited set of conditions in which the task is particularly challenging.

Author contributions

P.M.B. conceived the idea; W.J.H. and P.M.B. designed experiments; W.J.H. and J.M.V.M. performed the research; W.J.H. analysed data and wrote the initial draft, J.M.V.M. provided feedback and W.J.H. and P.M.B. revised the manuscript.

Supplementary data

All data are available on the Open Science Framework: <https://osf.io/vbhq/>

CRediT authorship contribution statement

William J. Harrison: Conceptualisation, Methodology, Formal Analysis, Writing – Original draft, Writing – Review and Editing, Visualisation, Investigation. J.M.V.M.: Writing – Review and Editing, Investigation. P.M.B.: Conceptualisation, Methodology, Formal Analysis, Resources, Writing – Review and Editing, Supervision.

Declaration of Competing Interest

None.

Acknowledgments

This work was supported by the Wellcome Trust (106926), the National Health and Medical Research Council of Australia (APP1091257), and the Australian Research Council (DE190100136). We thank Ms. Imogen Stead for her assistance collecting data for Experiment 6.

References

- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63(1), 1–29. <https://doi.org/10.1146/annurev-psych-120710-100422>.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12), 1–18. <https://doi.org/10.1167/9.12.13>.
- 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. <https://doi.org/10.1177/0956797610397956>.
- Brady, T. F., & Alvarez, G. A. (2015a). Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of Vision*, 15(15), 6. <https://doi.org/10.1167/15.15.6>.
- Brady, T. F., & Alvarez, G. A. (2015b). No evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 921–929. <https://doi.org/10.1037/xlm0000075>.
- Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, 138(4), 487–502. <https://doi.org/10.1037/a0016797>.
- Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, 120(1), 85–109. <https://doi.org/10.1037/a0030779>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, 20(5), 324–335. <https://doi.org/10.1016/j.tics.2016.03.006>.
- Corbett, J. E. (2017). The whole warps the sum of its parts. *Psychological Science*, 28(1), 12–22. <https://doi.org/10.1177/0956797616671524>.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The EyeLink toolbox: Eye tracking with MATLAB and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617.
- Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, 18(5), 1016. <https://doi.org/10.1364/JOSAA.18.001016>.
- Dakin, S. C., Cass, J., Greenwood, J. A., & Bex, P. J. (2010). Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *Journal of Vision*, 10(10), 14. <https://doi.org/10.1167/10.10.14>.
- Harrison, W. J., & Bex, P. J. (2015). A unifying model of orientation crowding in peripheral vision. *Current Biology*, 25(24), 3213–3219. <https://doi.org/10.1016/j.cub.2015.10.052>.
- Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, 75(2), 278–286. <https://doi.org/10.3758/s13414-012-0399-4>.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356. <https://doi.org/10.1038/nn.3655>.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. [https://doi.org/10.1016/0376-6357\(93\)90083-4](https://doi.org/10.1016/0376-6357(93)90083-4).
- Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, 15(4), 6–18. <https://doi.org/10.1167/15.4.6>.
- Morey, R. D. (2018). Package ‘BayesFactor’ (0.9.12-4.2) [Computer software]. <https://richarddmorey.github.io/BayesFactor/>.
- 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), 486–511. <https://doi.org/10.1037/rev0000101>.
- Orhan, A. E., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological Review*, 120(2), 297–328. <https://doi.org/10.1037/a0031541>.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, 18(4), 682–689. <https://doi.org/10.3758/s13423-011-0088-7>.
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, 10(14), 19. <https://doi.org/10.1167/10.14.19>.
- Utochkin, I. S., & Brady, T. F. (2020). Individual representations in visual working memory inherit ensemble properties. *Journal of Experimental Psychology: Human Perception and Performance*. <https://doi.org/10.1037/xhp0000727>.
- Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica*, 146, 7–18. <https://doi.org/10.1016/j.actpsy.2013.11.012>.
- Victor, J. D., & Conte, M. M. (2004). Visual working memory for image statistics. *Vision Research*, 44(6), 541–556. <https://doi.org/10.1016/j.visres.2003.11.001>.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217. <https://doi.org/10.1037/a0029333>.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 1120–1135. <https://doi.org/10.1167/4.12.11>.