Object-based encoding constrains storage in visual working memory

William X. Q. Ngiam[1], Krystian B. Loetscher[1], Edward Awh[1]

Affiliations: [1]Department of Psychology, University of Chicago

Correspondence should be addressed to William X. Q. Ngiam; E-mail: wngiam@uchicago.edu

Contributions: Conceptualization: W.X.Q.N. and E.A. Data curation: W.X.Q.N. Formal analysis: W.X.Q.N. Funding acquisition: E.A. Investigation: W.X.Q.N. and K.B.L. Methodology: W.X.Q.N. and E.A. Project administration: W.X.Q.N. Resources: W.X.Q.N. Software: W.X.Q.N. Validation: W.X.Q.N. Visualization: W.X.Q.N. Writing - original draft: W.X.Q.N. and E.A. Writing - review & editing: W.X.Q.N. and E.A.

Author Note

The data and ideas in the present manuscript were presented at the 2020 and 2022 Vision Sciences Society Meeting, the 2020 Annual Meeting of the Psychonomics Society, and the 2022 Object Perception, Attention and Memory conference. All experimental materials – task code, analysis code and experimental data – can be accessed at https://osf.io/wjr7u/ with DOI 10.17605/OSF.IO/WJR7U. The authors declare no conflict of interest. William X. Q. Ngiam and Edward Awh was supported by National Institutes of Health 2R01MH087214.

Abstract

The fundamental unit of visual working memory (WM) has been debated for decades. WM could be object-based, such that capacity is set by the number of individuated objects, or feature-based, such that capacity is determined by the total number of feature values stored. The present work examined whether object- or feature-based models would best explain how multi-feature objects (i.e., color/orientation or color/shape) are encoded into visual WM. If maximum capacity is limited by the number of individuated objects, then above-chance performance should be restricted to the same number of items as in a single feature condition. By contrast, if the capacity is determined by independent storage resources for distinct features – without respect to the objects that contain those features – then successful storage of feature values could be distributed across a larger number of objects than when only a single feature is relevant. We conducted four experiments using a whole-report task in which subjects reported both features from every item in a six-item array. The crucial finding was that above-chance recall – for both single- and multi-featured objects – was restricted to the first three or four responses, while the later responses were best modeled as guesses. Thus, whole-report with multi-feature objects reveals a distribution of recalled features that indicates an object-based limit on working memory capacity.

Public Significance Statement

Whether features or objects are the fundamental unit of visual working memory has been debated for decades. Across four experiments, we show object-based and feature-based effects occurring in concert – more feature values are remembered when memory is organized as objects (known as an object-based benefit), but features can be forgotten independently. Critically, we show that accurate recall of multi-feature objects is concentrated to the first three objects reported, contrary to feature-based accounts in which separate features are independently encoded without respect to the objects that contain them. Thus, object-based encoding limits constrain storage in visual working memory.

Introduction

Visual working memory (WM) is an *online* memory system responsible for maintaining information for ongoing cognition. Although its sharp capacity limitations are widely acknowledged, debate has persisted regarding the key limiting factors for storage in visual WM. On the one hand, *object-based* models propose that the limit is defined by a maximum number of objects that can be simultaneously held in visual WM (Luck & Vogel, 1997; Vogel et al., 2001). On the other hand, *feature-based* models propose that capacity is determined by the total amount of featural information stored (Wheeler & Treisman, 2002; Wilken & Ma, 2004). There was a surge of research on this issue following the seminal paper by Luck and Vogel (1997). They observed change-detection performance was unaffected by the addition of multiple features to visual stimuli in a WM task. Luck and Vogel argued that WM capacity was limited by the number of discrete objects encoded into memory, and not by the number of relevant features within each item. In line with this view, subsequent work has confirmed that more features can be stored when they are packaged within a smaller number of objects – an effect referred to as the *object-based benefit* (Fougnie et al., 2012; Gao et al., 2011; Hardman & Cowan, 2015; Lin et al., 2021; Olson & Jiang, 2002; Vogel et al., 2001).

However, it has also become clear that the perfect equivalence between WM performance with single- and multi-feature objects does not fully replicate. WM performance does decline as the number of features per item increases (Cowan et al., 2013; Hardman & Cowan, 2015; Markov et al., 2019; Oberauer & Eichenberger, 2013), although the size of the decline with multi-feature objects is far smaller than expected according to a pure feature-based model of capacity. For example, Hardman and Cowan (2015) in direct replications of Luck and Vogel (1997), consistently found that memory performance declined with additional feature load but noted that visual WM performance still depended on the total number of objects stored. Thus, it appears that WM performance cannot be predicted solely by the number of objects or features to be stored.

Here, we examined this issue by using a *whole-report* procedure in which recall for all features of every to-be-remembered item was tested during each trial (Adam et al., 2015, 2017; Hakim et al., 2020; Schor et al., 2020). The key virtue of this approach is that it provides a richer picture of the *distribution* of successfully stored features across the six items in the memory array. Observers in the Adam et al. (2015, 2017) studies saw displays of six colors or six orientations and were required to recall all colors or orientations. They showed a strong tendency to report the best-remembered items first, such that the first three responses showed clear evidence of target-related information, while the final three responses were best modeled as guesses (specifically, uniform distributions that were random with respect to the recalled items). Thus, these findings provided clear evidence that WM storage was restricted to about 3-4 items. In the present work, our goal was to leverage this whole-report procedure to examine the distribution of successfully stored feature values when there were multiple relevant features for each item. The key question was whether above-chance storage of relevant features would be constrained to a similar number of items with single- and multi-feature conditions, as predicted by object-based models of WM capacity.

We examined three models of WM capacity limits: a *strong object* model, an *independent features* model, and an *object-based pointer* model with feature loss. The *strong object* model presumes that capacity limits are determined entirely by the number of objects stored, and that all features of the stored objects are retained without loss (e.g.,Luck and Vogel, 1997; Olson & Jiang, 2002; Sone et al., 2021). The *independent features* model, by contrast, asserts that WM capacity is determined only by the total amount of feature information that is stored, without respect to the number of objects containing those features (e.g. Fougnie & Alvarez, 2011; Shin & Ma, 2017; Wheeler & Treisman, 2002). Finally, the *object-based pointer* model presumes that WM storage is constrained to a fixed number of objects, but also allows for independent forgetting of the features within the stored objects (e.g. Brady et al., 2011; Cowan et al., 2013; Markov et al., 2019; Oberauer & Eichenberger, 2013). To anticipate the results, we replicated past findings that subjects had a strong tendency to report the best remembered items first, such that above chance responses were restricted to the first three or four responses across the trial. The key finding was that this empirical pattern was observed for both single and multi-featured objects, suggesting that a similar item limit constrained performance for these stimuli. Thus, observers were not able to achieve above-chance performance with a larger number of multi-featured objects, posing a challenge to models that assert independent resources for the storage of distinct visual features. In line with this conclusion, formal model comparisons showed that the data were best fit by the *object-based pointer* model, providing evidence for *object-based encoding* into WM. Thus, our findings suggest that WM storage is limited by the number of objects stored, but features may be lost independently from those objects (Li et al., 2022; Sone et al., 2021).

Experiment 1 Methods

Transparency and Openness

All experimental materials – task code, analysis code and experimental data – can be accessed at https://osf.io/wjr7u/ with DOI 10.17605/OSF.IO/WJR7U. Data was analyzed using MATLAB 2019a (The MathWorks, Natick, MA) using custom analysis scripts. The study designs and their analysis were not pre-registered.

Apparatus

Stimuli were generated using MATLAB 2019a (The MathWorks, Natick, MA) and PsychToolbox (Brainard, 1997; Pelli, 1997), and presented on a 24-inch BenQ XL2430T LCD monitor with spatial resolution set to 1920 $\times$ 1080 and refresh rate set to 120 Hz. Participants were seated in a dark room with a viewing distance of approximately 80 cm.

Stimuli

Participants were shown displays containing six circular items with a diameter of 2.5° of visual angle in size. Items were centered within a region at least 2.5° of visual angle but less than 10° to either side of the central fixation point, with the constraint that three items were on both sides of the screen and items were at least 5° away from each other (see *Figure 1*).

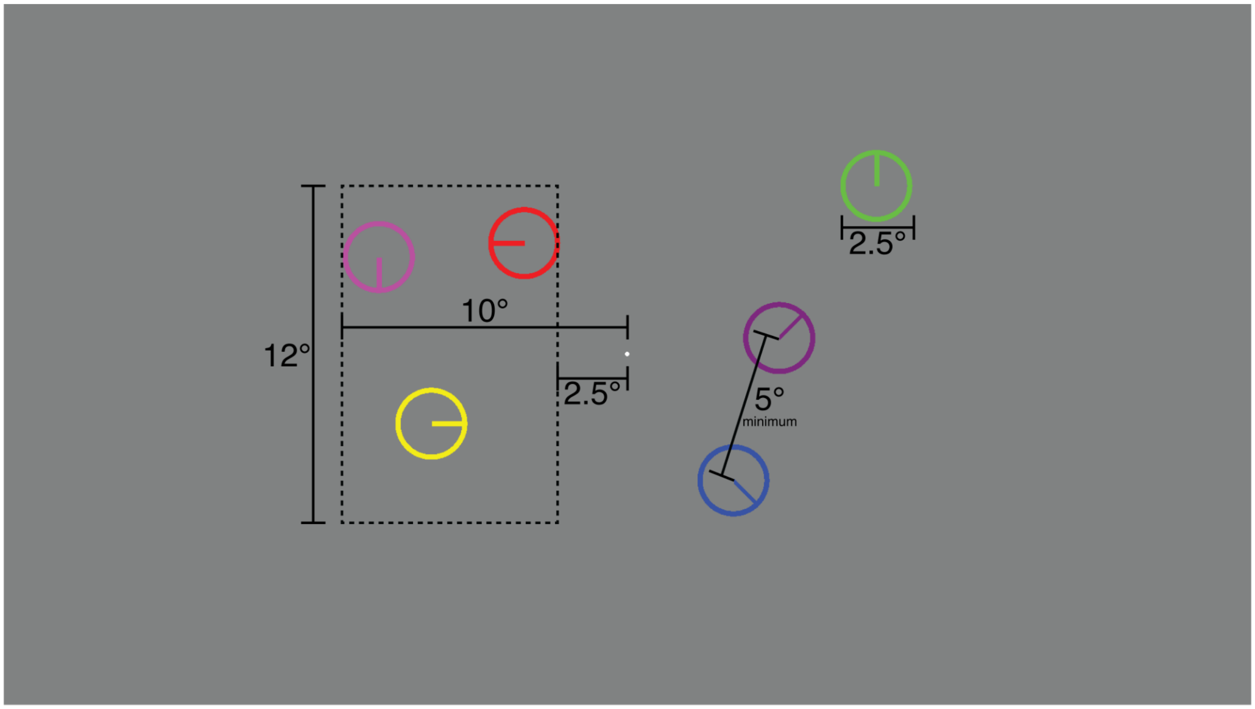


Figure 1. An example stimulus array with the constraints for item locations overlayed. Stimulus arrays always contained six items of 2.5° of visual angle in size, with three items to either side of the fixation point. Locations were randomly sampled from a region with a minimum distance from the fixation point was 2.5° of visual angle and the maximum distance was 10° of visual angle. The height of this region was 12° of visual angle. The items were separated by a minimum of 5°.

In Experiment 1, the displays were either colored circles, clock faces or colored clock faces (the conjunction of the single-feature conditions, see Figure 2). There were eight possible colors: red (RGB 255, 0, 0), green (0, 255, 0), blue (0, 0, 255), yellow (255, 255, 0), magenta (255, 0 , 255), cyan (0, 255, 255), orange (255, 128, 0), and purple (128, 0, 128). There were eight possible orientations (0°, 45°, 90°, 135°, 180°, 225°, 270° and 315° from vertical). In the orientation-only condition, the stimuli were presented in white. Only one of each color and orientation were randomly sampled and displayed in the memory array on each trial (see *Figure 2* for examples).

Figure 2. An example of stimulus arrays from each condition in Experiment 1 (orientation only, color only, and conjunction).

Procedure

The experiment was a discrete whole-report visual working memory task (see *Figure 3* for the general trial procedure). Each trial commenced with a fixation point displayed for 1000 msec before the memory display was shown for 500 msec. After a retention period of 1000 msec, placeholder circles appeared at the locations of each memory item on that trial. Participants were instructed to recreate the entire memory display using a click-and-drag response at each location – to respond, participants clicked within the placeholder they wished to respond to, and while holding down the mouse button, dragged outside the placeholder circle to select their response. In the color only condition, a wheel containing all possible colors appeared around the clicked location, and the observers moved their mouse cursor, a crosshair, to a set radius to select their desired color. In the orientation only condition, the observer dragged their mouse cursor in the direction of their desired response. In the conjunction condition, participants responded on both dimensions simultaneously – the observer would drag their mouse cursor in the direction of their desired orientation response, as well as to the color wheel of their desired color response. A preview of their selected response would appear within the placeholder circle, updating as they moved the cursor, allowing subjects to confirm their memory. Releasing the mouse button submitted the response. Participants were not allowed to change their responses once submitted, and could respond with the same color or orientation at multiple locations. The selected responses remained on screen until all items had a submitted response, after which the next trial started after a blank screen with an inter-trial interval of 1 second. Participants were given 10 practice trials of each condition for practice prior to the experiment.

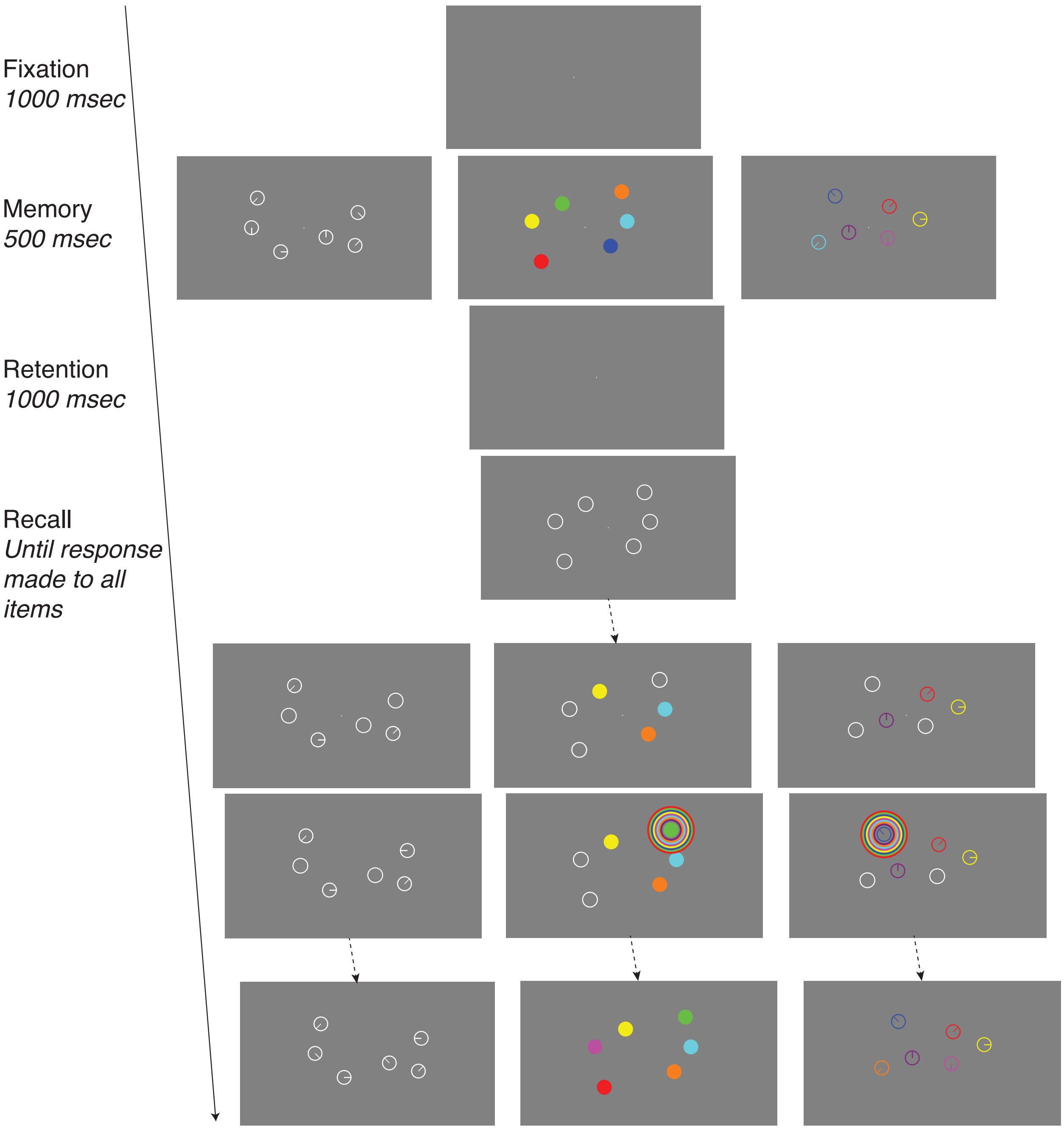


Figure 3. The general trial procedure. Each trial started a fixation point at the center of the screen displayed for 1000 msec. A memory array would then appear for 500 msec before a blank retention period of 1000 msec. The response screen starts with outlines of the locations of the memory items. Participants clicked within a location, and while holding down the button, dragged in a direction to indicate the orientation and/or dragged to a certain radius to indicate

the color, letting go to submit their response at that location. A preview was presented at the clicked location as they moved the cursor so the participant could confirm their response. Responded items remained on the screen until all responses were submitted.

Participants completed the experiment in two sessions, with the second session on a separate day but within 8 days of the first session. In each session, the participants completed 150 trials of each condition (color only, orientation only and conjunction), for a total of 900 trials across both sessions and all conditions. Participants were required to take at least a 30-second break every 50 trials. Condition order was counterbalanced across all participants. The duration of each session was approximately 2 to 2.5 hours.

Participants

Thirty participants (21 females and 9 males) between 18 and 32 years of age (mean age = 23.6 years) were recruited from the local University of Chicago community and received monetary compensation for their time across two sessions ($10 per hour, and a $10 bonus for completion of both sessions). Participants were asked for their gender and age by the experimenter, who recorded the answer in a free-response box on-screen. All participants reported normal or corrected-to-normal visual acuity, normal color vision and gave informed consent. Procedures were approved by the University of Chicago Institutional Review Board.

Analysis

Working memory performance was quantified as the number of correct responses on each trial, in each condition. To describe an individual's performance, we calculated the mean number of correct responses across trials in each condition (color only, orientation only, and conjunction). We also fit a computational model (see the *partial drop* model from Hakim et al., 2020; also Schor et al., 2020, 2023) to estimate the maximal working memory capacity (*Kmax*) and an attentional control factor (alpha) parameter for each individual (see "the *object-based pointer* model" for further details). The critical analysis examined the *distribution* of correctly recalled features across all items in the conjunction condition. To compare the three competing models, we conducted formal model comparisons on each observer's recall performance across the six responses in the conjunction condition.

<u>Modeling of Accuracy Across Responses in the Conjunction Condition</u>

The central goal of this experiment was to examine how accurately recalled features were distributed across items, enabling a test of object-based versus feature-based models of WM capacity. We conducted formal model comparisons of three models: the *strong object* model, *the independent features* model and the *object-based pointer* model (Figure 4). The *strong object* model assumes that people store a discrete number of objects in working memory, with perfect retention of all relevant features. There will be no representations for the remaining items in the memory array, resulting in chance performance due to guessing ("all-or-nothing"). The *independent features* model assumes that a limited number of feature values are stored, without any constraint on the number of objects containing those features. The *object-based pointer* model assumes that storage of multiple features is constrained to a maximum number of objects, while allowing for independent dropping (or encoding failure) of the features within those objects. Examples of the pattern of behavior predicted by each model can be seen in *Figure 4*. Note that these predictions are based on the assumption that observers will report the best remembered items first, as observed in past work using whole report procedures (Adam et al., 2015, 2017; Ngiam et al., 2022).

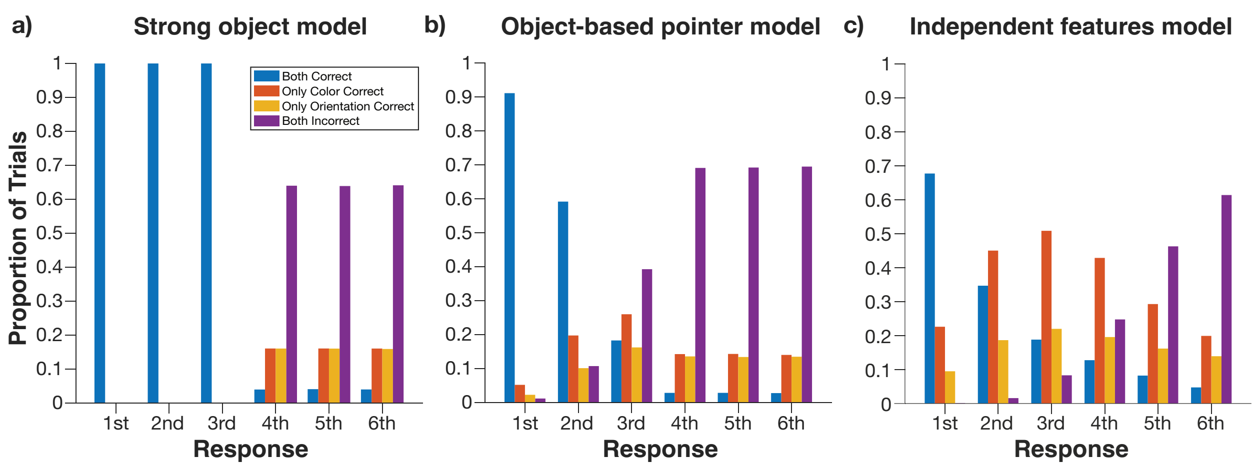


Figure 4. Example predictions for the pattern of responses for the compared models. (a) The

strong object model predicts memory representations are all-or-nothing. Depicted in this example is memory for 3 objects, and chance guessing for the remaining three responses. (b) The *object-based pointer* model predicts that working memory is defined by object-based encoding but may undergo feature loss. The rate of feature loss is independent for the separate feature dimensions. (c) The *independent features* model assumes success of encoding a feature into working memory is independent to the other feature of the object, and so working memory is distributed across objects. The rate of successful encoding may vary between the two feature dimensions.

We found the best-fitting parameter estimates for each of the three models depicted in *Figure 4* using likelihood-maximization procedures. We generated a probability distribution of recalling neither, just one of either feature, or both features of an object for each response (1st to 6th) according to each of the three models. We then determined the parameter estimates for each model that maximizes the likelihood for each individual participant's observed accuracy data.

<u>The strong object model</u>

The *strong object* model has one parameter, the number of objects remembered ($K \in [0, 6]$ with steps of 1). If the ordinal position of the response is less than or equal to this, the subject is assumed to correctly recall both features. For later responses, the object is assumed to make an informed guess, where the probability of the possible responses (both features correct, only one or the other feature correct, or both incorrect) is straightforwardly modeled with a binomial probability mass function. This is formulated as:

$$f(x,i) = \begin{cases} i \leq K, \begin{cases} p(both\ correct) = 1 \\ p(F_1\ correct) = 0 \\ p(F_2\ correct) = 0 \\ p(both\ incorrect) = 0 \end{cases} \\ i > K, \begin{cases} p(both\ correct) = \left(\frac{1}{8-K}\right)^2 \\ p(only\ F_1 correct) = \frac{7-K}{(8-K)^2} \\ p(only\ F_2\ correct) = \frac{7-K}{(8-K)^2} \\ p(both\ incorrect) = \left(\frac{1}{8-K}\right)^2 \end{cases} \end{cases},$$

where *x* is the type of response made, *i* is the ordinal position of the response, *F1* is the first feature dimension and *F2* is the second feature dimension. For each individual, we determined the best-fitting *strong object* model by calculating the maximum log-likelihood for their pattern of responses with each possible parameter value (K). The probability distribution when K equals 3 is shown as an example in Figure 4a).

The object-based pointer model

The *object-based pointer* model has four free parameters, the maximum number of objects remembered ($Kmax \in [1,6]$ with a step size of 1), the robustness of attentional control ($\alpha \in [0,10]$), and the probability of feature loss in each dimension (p_{loss_F1}; p_{loss_F2}). To determine the best-fitting parameter values for each individual, we used approximate Bayesian optimization to search for likely values that minimized the negative log-likelihood. To calculate the negative log-likelihood, we generated probability distributions at the specific combinations of parameter values via simulation (i.e. synthetic probability distributions). We simulated the number of objects successfully encoded based on a beta-binomial distribution previously demonstrated to explain performance on a working memory whole-report task (see *partial drop* model in Hakim et al. (2020) and also in Schor et al. (2020, 2023)). The beta-binomial distribution estimates the probability of successfully encoding an item from 0 to an individual's maximum capacity, *Kmax*. The α parameter of the beta function, $B(\alpha, \beta)$, shapes the graded distribution of the number of successfully stored items, capped by *Kmax* – the higher the α parameter value, the more often that *Kmax* (and close to *Kmax*) is achieved. This parameter has been shown to capture the fluctuations in attention in a working memory task following short and long retention intervals (Hakim et al., 2020). We then simulated the independent loss of each feature from each of the successfully encoded objects, as previously observed by Fougnie and Alvarez (2011). For the two feature dimensions, we estimated the number of times that a feature from a dimension was dropped from a successfully encoded object with a binomial distribution based on the probability of feature loss in that dimension (p_{loss_F1} , p_{loss_F2}) as a free parameter. It is important to note that these probabilities can vary for the two feature dimensions – that one feature may be lost more often than the other feature:

$$f(x; n_i) = x \sim Binom\left(n_i, p_{loss_{F1}}\right),$$

where x is the number of drops, n is the number of objects successfully encoded on the ith trial, and p_{loss_F1} is the probability of a feature from that dimension being lost. Following failures of encoding and independent feature loss, the subject is assumed to make an informed guess by choosing from the remaining feature values that were not successfully encoded. Observers are assumed to order responses based on the number of features they accurately recalled, as previously observed by Adam et al. (2017) and Ngiam et al. (2022). The synthetic probability distribution was estimated by taking the average distribution across 100 iterations simulating 500 trials.

We conducted maximum likelihood estimation by using the approximate Bayesian optimization algorithm in MATLAB (*bayesopt*). Bayesian optimization adapts the search by modeling the objective function with a Gaussian process model (in this case, the negative log-likelihood function across the four-parameter space). It can then balance testing underexplored areas in the parameter space and the parameter estimates that are more likely to generate the objective minimum. As we estimated the negative log-likelihood using synthetic probability distributions, the noise level of the objective function (the negative log-likelihood calculation) was also estimated during the minimization search. The approximate Bayesian optimization process was fixed to end after 100 evaluations.

The *object-based pointer* model is very similar in application to a model proposed by Cowan et al. (2013), whereby a constant number of objects is held in WM, and within each remembered object, one feature is stored with certainty but the success of storing remaining additional features is probabilistic (see Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013). Here, we have extended these mechanisms to generate probability distributions on the whole-report task with conjunction stimuli. An example of a probability distribution can be seen in Figure 4b).

The independent features model

The *independent features* model has two free parameters, the probability of successful encoding for each feature dimension (p_{F1}; p_{F2}). We generated probability distributions through simulation (a synthetic probability distribution) over the 2D grid space for both parameters ($p_{F1} \in$ [0.01,1] and $p_{F2} \in$ [0.01,1] with a step size of 0.01. To do this, we conducted 100 iterations simulating 500 whole-report trials and averaged across the iterations to generate the final probability distribution. On each trial, we simulated the successful encoding of each feature within each item using a Bernoulli distribution for each of the probability parameters for the two feature dimensions:

$$f(x, i) \sim Bernoulli(p_{F1}),$$

where x is whether the feature was successfully encoded for the ith item on the trial, and p_{F1} is the probability of successful encoding for that feature dimension. Like the object-based pointer model, on items where the encoding failed, we assumed informed guessing by simulating responses being chosen from the remaining possible feature values. Observers were assumed to order responses based on the number of features they accurately recalled, as previously observed by Adam et al. (2017) and Ngiam et al. (2022). An example of a distribution can be seen in Figure 4c).

To facilitate model comparison across the models with different numbers of estimated parameters, we converted the likelihood values to Bayes Information Criterion (BIC) values, which applies a penalty for additional model parameters (Schwarz, 1978). We used the lowest BIC value to select the best-fitting model (Kass & Raftery, 1995).

Experiment 1 Results

Accuracy Across Conditions

In the color-only condition, participants responded correctly to 3.21 items (SD = 0.74) on average per trial. In the orientation-only condition, the average number of orientations correctly reported per trial was 2.79 items (SD = 0.44). In the conjunction condition, the average number of correctly reported items (both features of the conjunction reported accurately) per trial was 1.62 items (SD = 0.38). The average number of accurately recalled features per trial in the conjunction condition was 4.94 features (SD = 0.68). The frequency distribution for number of correct responses for each condition can be seen in *Figure 5*.

In the single-feature conditions, we also fit a computational model to each individual's distribution of correct responses using maximum likelihood procedures, to estimate the maximal working memory capacity (*Kmax*) and attentional control factor (*alpha*) (Hakim et al., 2020; Schor et al., 2020, 2023). In the color-only condition, mean *Kmax* was 3.57 items (SD = 1.04) and mean *alpha* was 3.29 (SD = 1.93). In the orientation-only condition, mean *Kmax* was 2.70 items (SD = 0.47) and mean *alpha* was 4.94 (SD = 2.95).

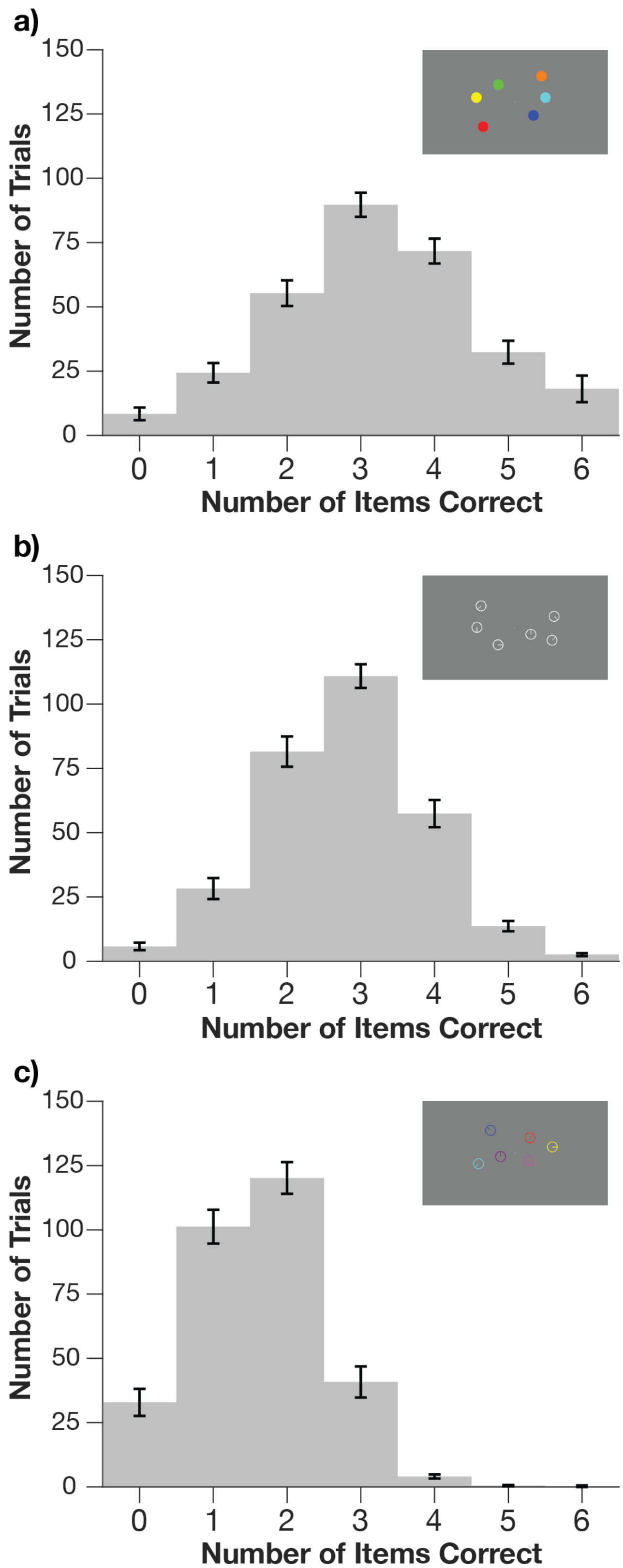
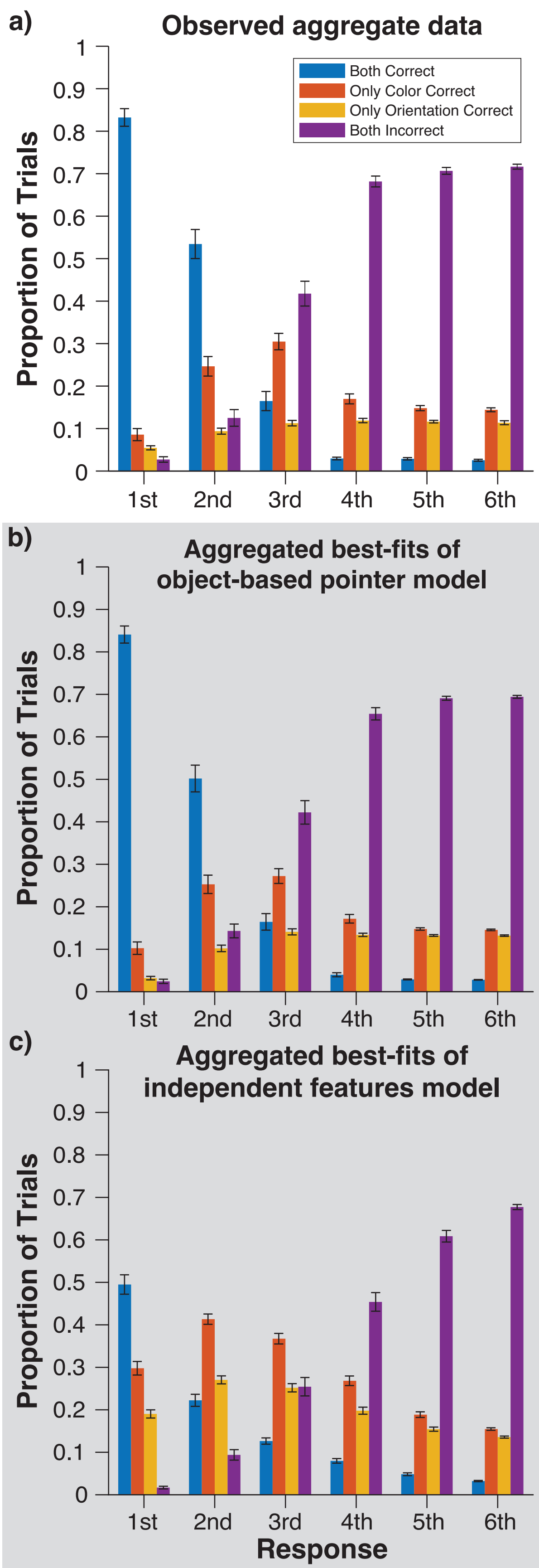


Figure 5. The average frequency histogram of the number of correct responses on each trial for (a) the color-only condition, (b) the orientation-only condition and (c) the conjunction condition of

Experiment 1. For the conjunction condition, a correct response was defined as reporting both features correctly. All error bars depict ±1 standard error of the mean.

Accuracy Across Responses in the Conjunction Condition

We then examined the distribution of accuracy across responses in the conjunction condition. Recall performance worsened across responses (*Figure 6a*). In the first three responses, participants reported at least one feature accurately more often than not. In the last three responses, however, the pattern of responding was near-identical and resembled chance performance. That is, accurate recall was constrained to the first three objects reported. We conducted formal model comparisons to verify our visual inspection of this empirical pattern.



a)
Observed aggregate data
Both Correct
Only Color Correct
Only Orientation Correct
Both Incorrect
Proportion of Trials
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1st
2nd
3rd
4th
5th
6th
b)
Aggregated best-fits of object-based pointer model
Proportion of Trials
c)
Aggregated best-fits of independent features model
Proportion of Trials
Response

1

Figure 6. a) Average distribution of recall accuracy across responses observed across participants in Experiment 1. Each bar indicates a possible type of response: both features of the item correct (blue), only the color of the item correct (orange), only the orientation of the item correct (yellow), or both features reported incorrectly (purple). The error bars indicate ±1 standard error of the mean. The aggregated distribution generated from the best-fitting parameters for each participant b) for the object-based pointer model and c) for the independent features model are shown in the grey panels. The object-based pointer model was the best-fitting for all participants in Experiment 1.

<u>Model Comparison Results</u>

For all 30 participants, the best-fitting model was the *object-based pointer* model with substantial evidence (Figure 6b) (mean ΔBIC to the *strong object* model = 5910.8 and mean ΔBIC to the *independent features* model = 494.10, see Figure 6c). Taking the best-fitting parameter estimates of the *object-based pointer* model for each individual, the average number of 'pointers' (*Kmax*) was 3.43 (SD = 0.97), the average *alpha* parameter was 6.62 (SD = 2.53), the probability of feature loss for color was 0.27 (SD = 0.14) and the probability of feature loss for orientation was 0.41 (SD = 0.20). Of note, the *Kmax* parameter estimate (the maximal number of items stored) was comparable between the single-feature and conjunction conditions.

Experiment 1 Discussion

We replicated past work showing that there is a consistent cost for WM performance with conjunction compared to single feature items (Hardman & Cowan, 2015; Markov et al., 2019; Olson & Jiang, 2002). We also observed an *object-based benefit* (Fougnie et al., 2012) whereby more features were accurately recalled in the conjunction condition (~5 features) compared to the single-feature conditions (~3 features). The key novel contribution of this work, however, was to leverage the whole report procedure to examine how the recalled features were distributed across items in the single- and multi-feature object conditions. If WM capacity is limited by the number of objects stored, then accurate recall with multi-feature objects should be constrained to the same number of items as with single-feature objects. Alternatively, if there are independent resources for storing distinct features, without regard for the objects that contain them, then correctly recalled features could be distributed across a larger number of objects in the multi-feature compared to the single-feature condition.

Recall performance worsened across the responses, as in past work (Adam et al., 2017; Ngiam et al., 2022). Accurate recall was restricted to early responses, whereas the later responses resembled chance performance. Our formal model comparison confirmed that the recall of these features was concentrated to within 3 items rather than distributed across a larger number of items in the display. Critically, this was the case for both single- and multi-feature objects. For all 30 participants, the *object-based pointer* model provided a better fit to the data than the *independent features* model, which assumes no object-based constraints on the distribution of featural memory. Note that the *object-based* model also outperformed the *strong object* model, suggesting that object-based storage is not lossless, as features are lost independently from each object memory representation. Thus, our findings indicate that WM storage is constrained by the number of objects stored.

Experiment 2

As recall accuracy was significantly better for color than orientation in the single-feature condition, we were concerned VWM accuracy could have been inflated due to color grouping strategies. Indeed, debriefing of Experiment 1 participants revealed that some made active use of grouping strategies based on “warm” and cool” color categories. This motivated a replication of Experiment 1 with a shorter presentation time of the memory array (150 msec in this experiment compared to 500 msec in Experiment 1) to reduce the opportunity for these strategies.

Experiment 2 Methods

Participants

30 observers (14 females and 16 males) between 19 and 35 years of age (mean age = 23.9) who did not complete the previous experiment received the same monetary compensation

for their participation across two sessions. Each session took approximately two hours and the second session was completed on a separate day within 11 days of the first session.

<u>Procedure</u>

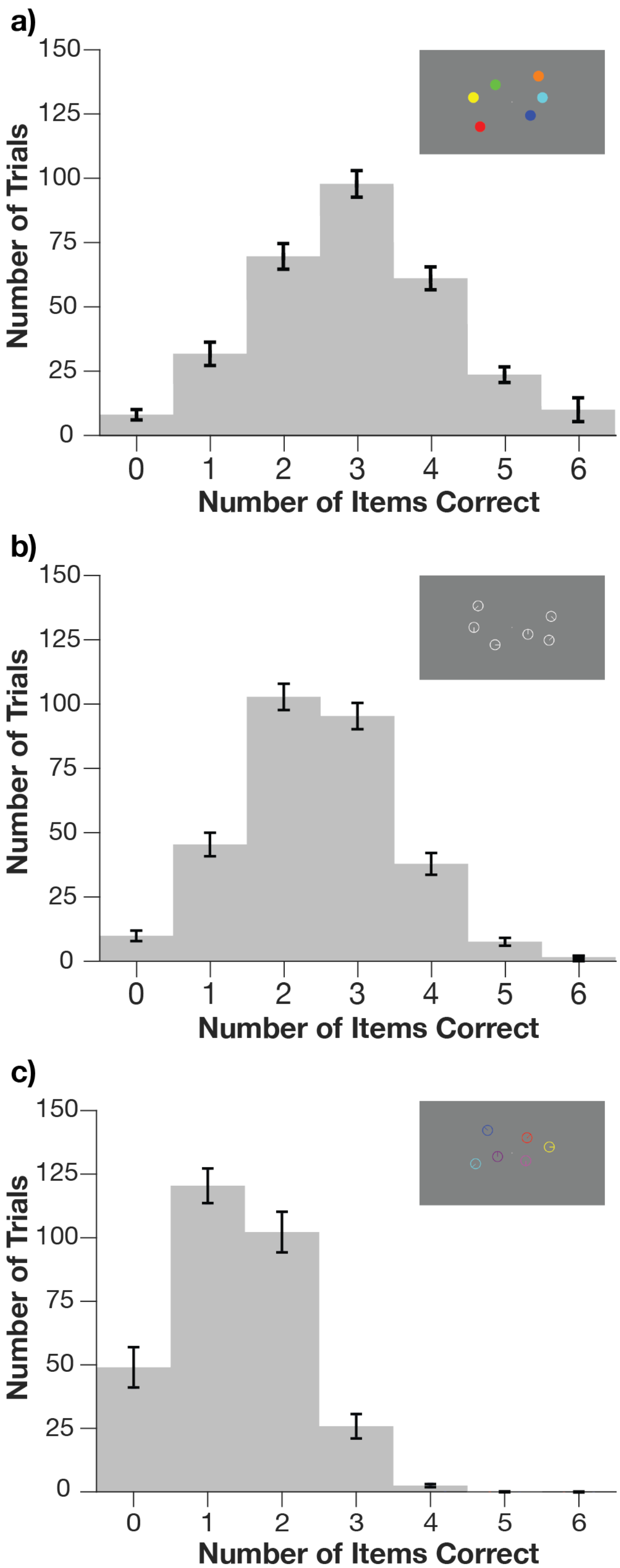
The procedure was identical to Experiment 1 except that the memory array was presented for 150 msec rather than 500 msec.

Experiment 2 Results

<u>Accuracy</u>

The mean number of colors recalled per trial was 2.94 items (SD = 0.64). The mean number of orientations recalled per trial was 2.45 items (SD = 0.45). The mean number of conjunctions correctly reported per trial was 1.38 items (SD = 0.42), and the mean number of features in the conjunction condition correctly reported was 4.52 features (SD = 0.83). The aggregated distribution of correct responses for each condition is displayed in *Figure 7*.

We fit a computational model to each individual's distribution of correct responses to estimate their maximal capacity (*Kmax*) and their attentional control (*alpha*). In the color-only condition, mean *Kmax* was 3.17 (SD = 0.79) and mean *alpha* was 3.17 (SD = 1.81). In the orientation-only condition mean *Kmax* was 2.40 (SD = 0.56) and mean *alpha* was 3.92 (SD = 2.60).


a)
Number of Trials
Number of Items Correct
b)
Number of Trials
Number of Items Correct
c)
Number of Trials
Number of Items Correct

1

Figure 7. The average frequency histogram of the number of correct responses on each trial for (a) the color-only condition, (b) the orientation-only condition and (c) the conjunction condition of Experiment 2. All error bars depict ±1 standard error of the mean.

<u>Accuracy Across Responses in the Conjunction Condition</u>

The pattern of results was very similar to Experiment 1 – recall accuracy worsened across responses (see *Figure 8a*). Again, above-chance recall was constrained to the first three responses, whereas the last three responses resembled chance performance. Model comparison using likelihood-maximization procedures revealed that for all 30 participants, the best-fitting model was the *object-based pointer* model with substantial evidence (see Figure 8b) (mean ΔBIC to the *strong object* model = 6617.7 and mean ΔBIC to the *independent features* model = 415.31, see Figure 8c). Taking the best-fitting parameter estimates of the *object-based pointer* model for each individual, the average number of 'pointers' (*Kmax*) was 3.23 (SD = 1.19), the average alpha parameter was 6.57 (SD = 2.48), the probability of feature loss for color was 0.31 (SD = 0.14), and the probability of feature loss for orientation was 0.47 (SD = 0.19). Again, estimates of *Kmax* from the conjunction condition were comparable to the estimates in the single-feature conditions.

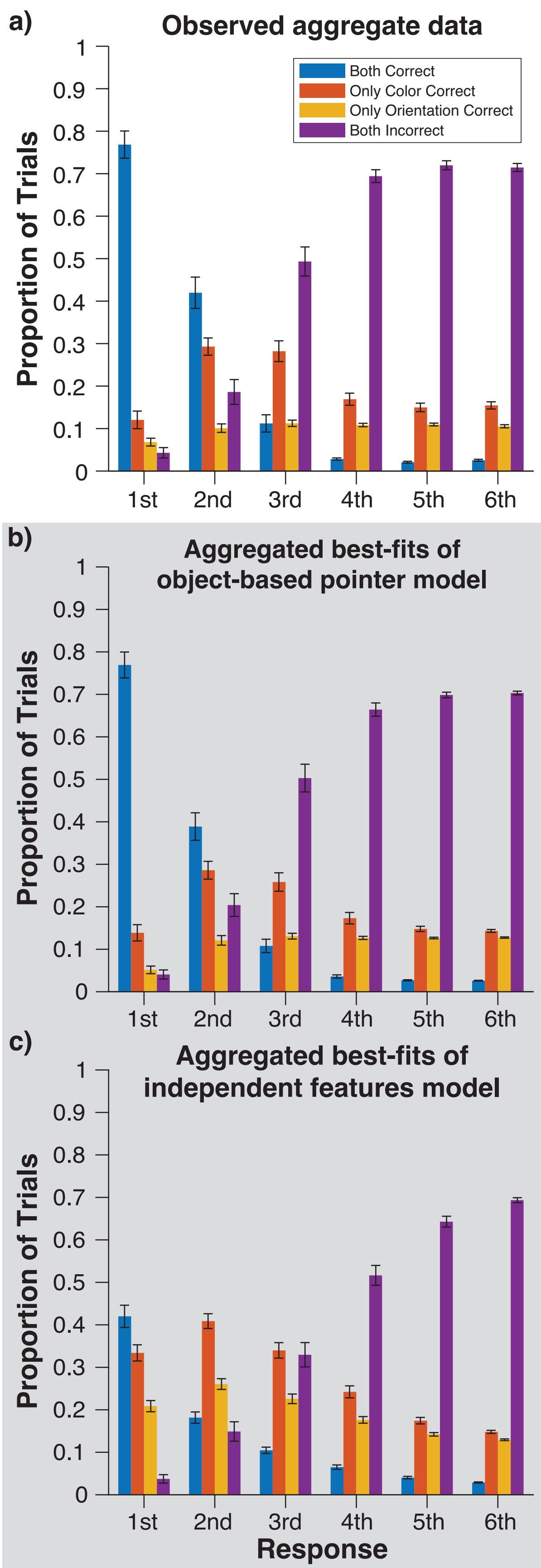

a)
Observed aggregate data
Both Correct
Only Color Correct
Only Orientation Correct
Both Incorrect
Proportion of Trials
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1st
2nd
3rd
4th
5th
6th
b)
Aggregated best-fits of
object-based pointer model
Proportion of Trials
1st
2nd
3rd
4th
5th
6th
c)
Aggregated best-fits of
independent features model
Proportion of Trials
1st
2nd
3rd
4th
5th
6th
Response

Figure 8. a) Memory recall for conjunctions of color and orientation on each response in Experiment 2, averaged across participants. Error bars represent ±1 standard error of the mean. Displayed in the grey panel is the average of the recall distributions generated from the best-fitting parameters to each individual participant's data for the b) object-based pointer model and for the c) independent features model. The object-based pointer model was best-fitting for all 30 participants.

Experiment 2 Discussion

Like in Experiment 1, we observed that fewer conjunctions were remembered overall compared to single-feature items. Again, we observed the *object-based benefit* whereby more features were recalled overall in the conjunction condition (4.52 features) compared to the single-feature conditions (2.94 colors and 2.45 angles). Most notably, we replicated the key finding that the distribution of accurate recall in the conjunction condition was again concentrated within the first three responses, as confirmed by formal model comparisons – the *object-based pointer* model was best-fitting for all 30 observers compared to the *strong object* model and the *independent features* model. Therefore, it appears that while *object-based* storage allows more feature values to be stored from multi-featured compared to single-feature objects, there is probabilistic loss (or failure to encode) the features within each object.

Experiment 3

In Experiment 3, we directly replicated the conjunction condition of Experiment 1, except that we used another set of stimuli typically used in the literature to examine memory for conjunctions of angle and color – colored triangles. This would allow us to better compare results to previous experiments (Fougnie & Alvarez, 2011).

Experiment 3 Methods

The method for Experiment 3 was identical to Experiment 1 except for the following:

<u>Participants</u>

30 participants (21 females and 9 males) between the ages of 18 and 35 years (mean age = 24.47) completed this experiment. The participants had not completed any of the previous

experiments. Participants completed all trials in a single session, receiving the same rate of monetary compensation. The experimental session took approximately 2 hours to complete.

Stimuli

Instead of the colored clock faces, we presented colored triangles – another stimulus that has been used to examine memory for the conjunction of color and angle. The triangles had internal angles of 35°, 75° and 75° with side length of 2.5°.

Procedure

This experiment did not include the single-feature conditions – participants completed 300 trials of the conjunction condition only.

Experiment 3 Results

Accuracy

The mean number of conjunctions correctly reported per trial was 1.47 items (SD = 0.44), and the mean number of features correctly reported per trial was 5.11 (SD = 0.65) (see *Figure 9* for the aggregate distribution).

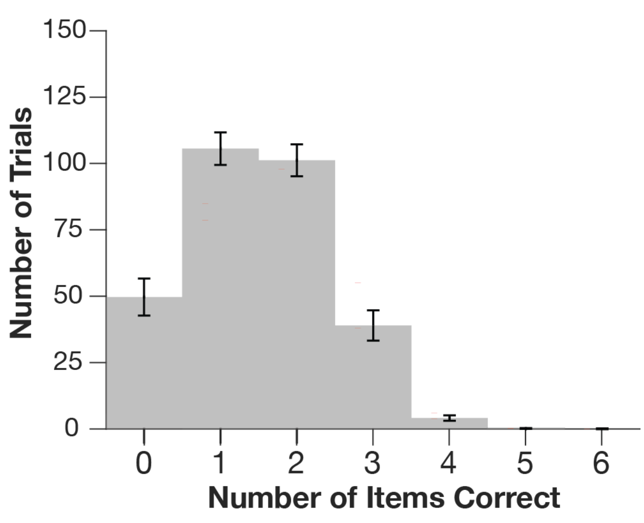
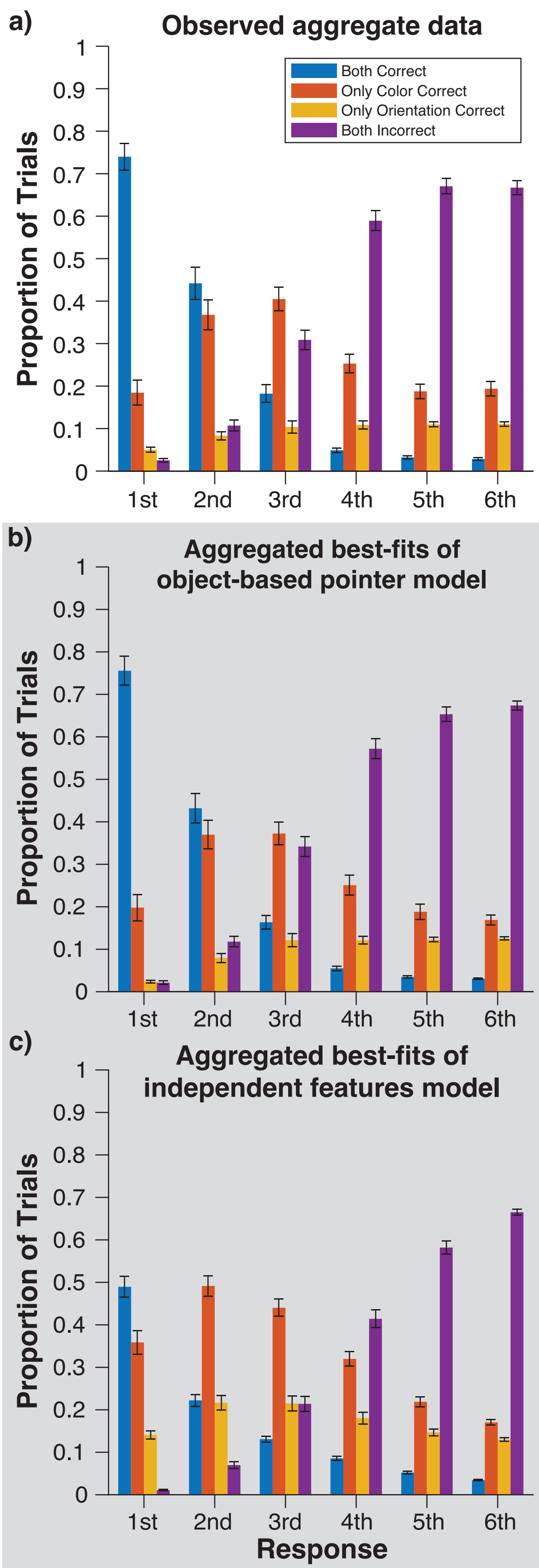


Figure 9. The average accuracy distribution of conjunctions remembered in Experiment 3. All error bars depict ±1 standard error of the mean.

Model Comparison

We observed the same pattern of results in Experiment 1 and 2 in the conjunction condition across responses (see *Figure 10a*). For all 30 participants, the best-fitting model was the *object-based pointer* model (see Figure 10b) (mean ΔBIC to the *strong object* model = 5997.4 and mean ΔBIC to the *independent features* model = 326.33, see Figure 10c). Taking the best-fitting parameters of the *object-based pointer* model for each individual, the mean number of 'pointers' (*Kmax*) was 4.30 (SD = 1.15), the mean alpha parameter was 6.16 (SD = 2.80), the mean probability of color loss was 0.29 (SD = 0.17) and the mean probability of feature loss for orientation was 0.58 (SD = 0.23).


a)
Observed aggregate data
Both Correct
Only Color Correct
Only Orientation Correct
Both Incorrect
Proportion of Trials
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1st
2nd
3rd
4th
5th
6th
b)
Aggregated best-fits of
object-based pointer model
Proportion of Trials
1st
2nd
3rd
4th
5th
6th
c)
Aggregated best-fits of
independent features model
Proportion of Trials
1st
2nd
3rd
4th
5th
6th
Response

1

Figure 10. Average memory recall accuracy for each response for Experiment 3. Model-fitting results are presented in the grey panels: b) Average distribution of recall accuracy generated by the best-fitting object-based pointer model for each individual. c) Average distribution of the recall accuracy distribution generated by the best-fitting independent features model for each individual. The object-based pointer model was the best-fitting for all 30 participants' data. The error bars depict ±1 standard error of the mean.

Experiment 3 Discussion

We replicated the results from the conjunction condition of our earlier experiments – accurate recall was concentrated to the first three responses with chance performance for the remaining three responses, and that the *object-based pointer* model better fit every participants' data compared to the *strong object* model and the *independent features* model. The improved recall accuracy for color over orientation appears to be widened with colored isosceles triangles, compared to the colored clock faces from Experiment 1 and Experiment 2. Our impression is that the orientations were more difficult to discriminate with these stimuli.

Experiment 4

Having consistently observed a recall advantage for color over orientation across the first three experiments, we sought to replicate the previous experiments while attempting to equate memory performance for the different feature dimensions. We used shapes rather than orientations and selected a new set of colors. For both features, our aim was to maximize the discriminability of every possible pair of values for each set. In addition to providing a fourth replication of our primary empirical pattern, Experiment 4 also served to generalize our findings to a different pair of features.

Experiment 4 Methods

The method for Experiment 4 was identical to Experiment 1 except for the following:

Participants

30 participants (18 females, 2 non-binary and 10 males) between the ages of 19 and 33 years (mean age = 24.50) completed this experiment. These participants had not completed

any of the previous experiments. Each session took approximately two and a half hours and the second session was completed on a separate day within 7 days of the first session.

Stimuli

We selected a set of colors and shapes with the aim of maximizing discriminability between all possible pairs of items and thus, equating the memory performance across the feature dimensions. The colors (RGB values in brackets) were red [175,0,0], green [0,140,0], blue [0,0,255], yellow [255,255,0], magenta [255,0,255], white [255,255,255], orange [255,128,0] and black [0,0,0]. The shapes used can be seen in *Figure 11*.

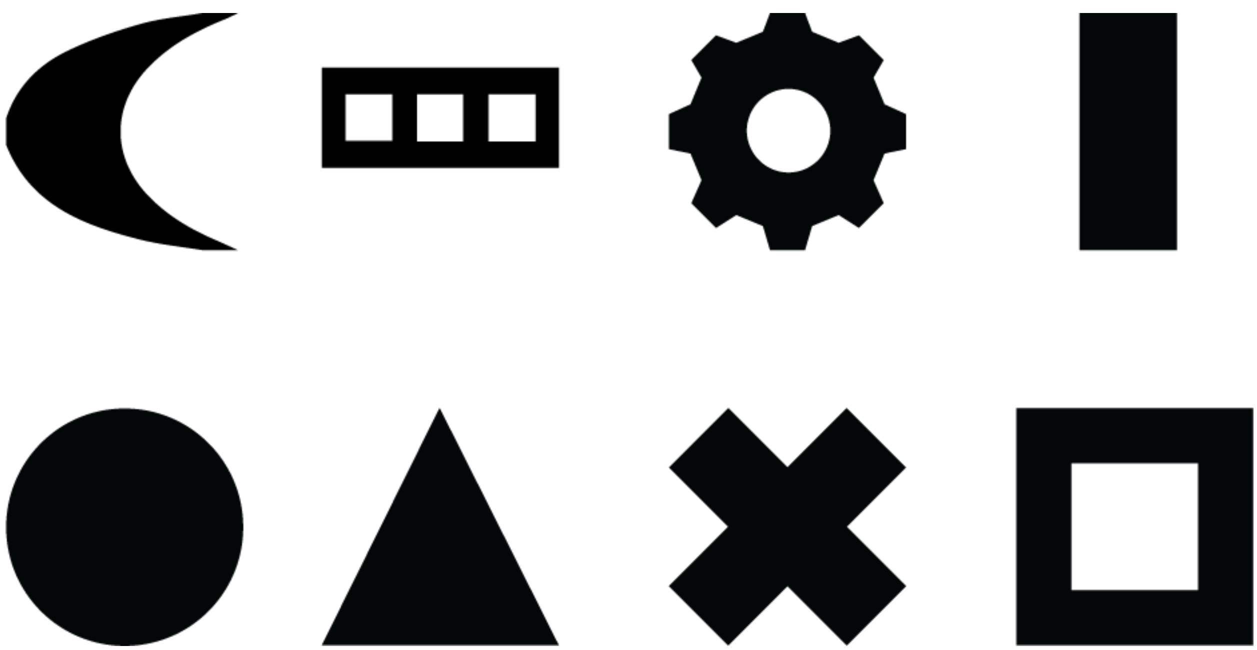


Figure 11. The discrete shapes used as stimuli in Experiment 4.

Procedure

The response screen for the shape-only and conjunction conditions included a cue of each shape around the perimeter of the response circle to indicate to the participant which direction to drag their cursor for their intended shape responses. Participants completed 5 practice trials of each condition (color-only, shape-only and the conjunction condition) to get used to the response interface and the experimental task.

Experiment 4 Results

Accuracy

In the single-feature conditions, the mean number of correctly recalled colors per trial was 3.61 items (SD = 0.75) and the mean number of correctly recalled shapes per trial was 3.39 items (SD = 0.64). The mean number of correctly recalled conjunctions (both shape and color features in the item) was 1.92 items (SD = 0.43), and the mean number of features recalled per trial of the conjunction condition was 5.34 features (SD = 0.85). The aggregated frequency distributions of correct responses for each condition are displayed in *Figure 12*.

We fit a computational model to the frequency distribution of correct responses in the single-feature conditions for each individual to estimate their maximal working memory capacity (*Kmax*) and attentional control (*alpha*). In the color-only condition, mean *Kmax* was 4.00 items (SD = 0.98) and mean *alpha* was 4.03 (SD = 2.44). In the shape-only condition, mean *Kmax* was 3.50 items (SD = 0.94) and mean *alpha* was 4.65 (SD = 2.56).

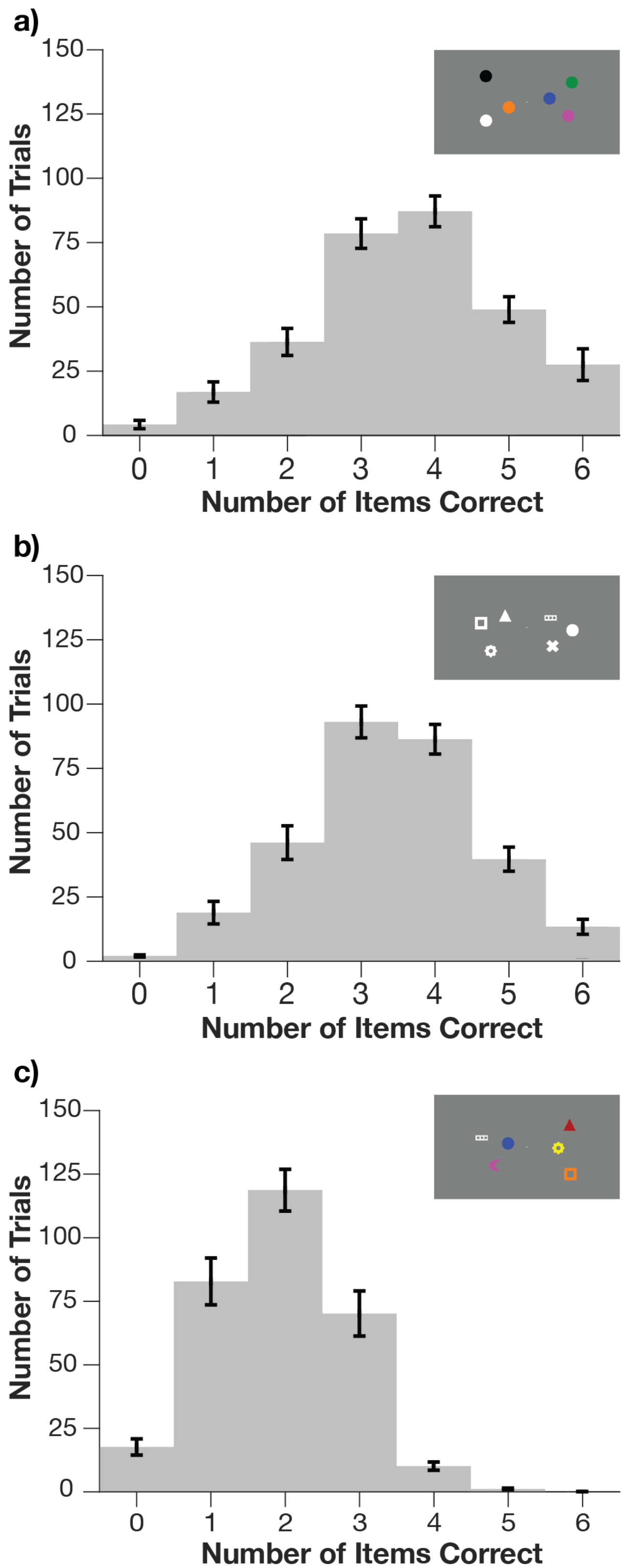

a)
Number of Trials
150
125
100
75
50
25
0
0 1 2 3 4 5 6
Number of Items Correct
b)
Number of Trials
150
125
100
75
50
25
0
0 1 2 3 4 5 6
Number of Items Correct
c)
Number of Trials
150
125
100
75
50
25
0
0 1 2 3 4 5 6
Number of Items Correct

Figure 12. The average frequency distribution of the number of correct responses in the (a) color-only condition (b) shape-only condition and (c) conjunction condition of Experiment 4. All error bars depict ±1 standard error of the mean.

Model Comparison

We observed the same pattern of results across responses in the conjunction condition as in the previous experiments – accurate recall was constrained to the first three responses, and the last three responses resembled chance performance (see *Figure 13a*). Of note, there did not appear to be a substantial advantage for one feature dimension over the other – color and shape were well-matched in recall accuracy across the responses. For a fourth time, the *object-based pointer* model was the best fitting model for all 30 participants (see Figure 13b) (mean ΔBIC to the *strong object* model = 5318.0 and mean ΔBIC to the *independent feature* model = 659.55 (see Figure 13c)). Taking the best-fitting parameters of the *object-based pointer* model for each individual, the mean number of ‘pointers’ (*Kmax*) was 3.20 (SD = 0.89), mean alpha was 6.10 (SD = 2.46), the mean probability of color loss was 0.24 (SD = 0.13) and the mean probability of shape loss was 0.23 (SD = 0.12).

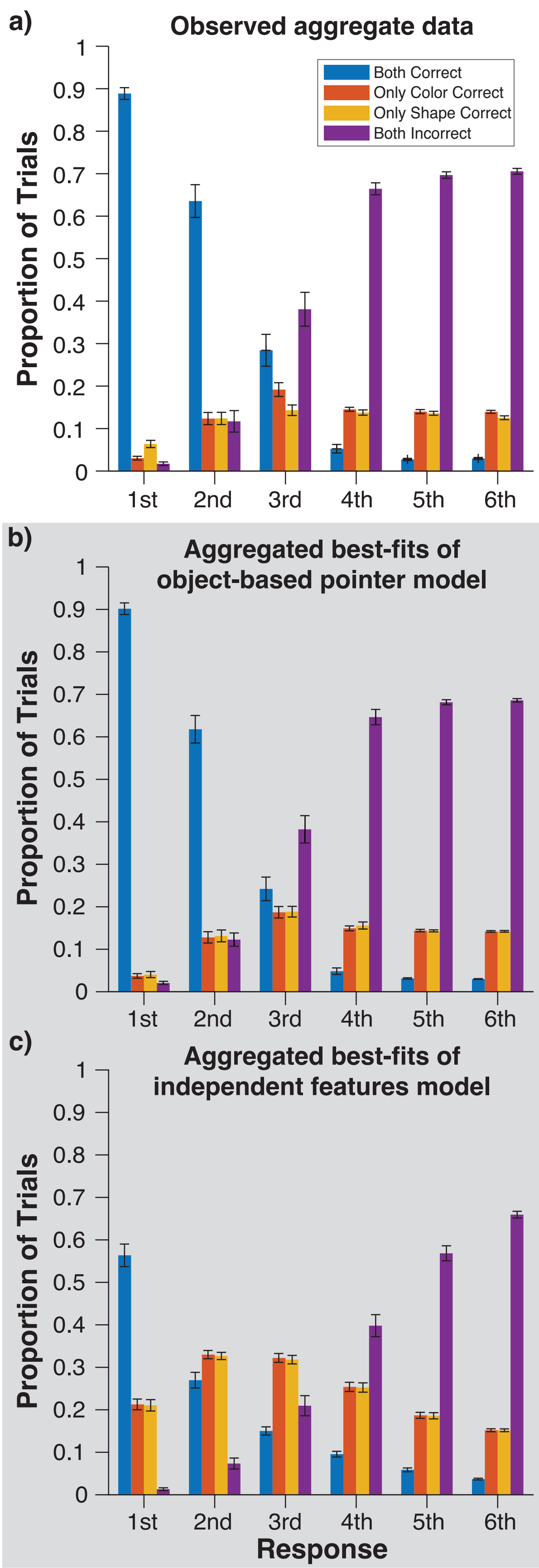

a)
Observed aggregate data
Both Correct
Only Color Correct
Only Shape Correct
Both Incorrect
Proportion of Trials
1st
2nd
3rd
4th
5th
6th
b)
Aggregated best-fits of object-based pointer model
c)
Aggregated best-fits of independent features model
Response

Figure 13. a) The observed average memory recall accuracy for each response for Experiment 4. Model-fitting results are presented in the grey panels: b) Average of the recall accuracy distribution generated by the best-fitting object-based pointer model for each individual. c) Average distribution of the recall accuracy distribution generated by the best-fitting independent features model for each individual. The error bars depict ±1 standard error of the mean.

Experiment 4 Discussion

In Experiment 4, we employed highly discriminable colors and shapes to equate memory performance across the stored features. In addition, this study generalized the findings from Experiments 1-3 to a new pair of features. As in the previous experiments, less conjunctions were perfectly remembered compared to the single-feature items. But again, we observed an *object-based benefit* – more features were accurately recalled in the conjunction condition (~5 features) compared to the single-feature conditions (~3 features). Notably, recall for color and shape appeared to be comparable in the single-feature conditions, as well as across responses in the conjunction condition. For a fourth time, we replicated the key finding from our earlier experiments – accurate recall was constrained to the first three responses, with guessing performance in later responses of the whole-report. This was confirmed by formal model comparisons that indicated that every participant's data was better fit by the *object-based pointer* model compared to the *strong object* model and the *independent features* model.

General Discussion

There has been persistent debate about whether capacity limits of visual WM are object-based or feature-based. The primary empirical approach within this debate has been to compare WM performance with single- and multi-feature stimuli, observing whether adding features to the memorized items impairs performance, as predicted by feature-based models of capacity. Although it is clear that adding additional features yields a measurable decline (e.g., Hardman & Cowan, 2015; Olson & Jiang, 2002), the effect is much smaller than it should be according to a pure feature load account of WM capacity. Instead, observers are able to store a substantially larger number of feature values when they are presented within multi-feature stimuli compared to single feature stimuli. In the present work, we replicated this well-known object-based benefit (Fougnie et al., 2012), disconfirming pure feature-based models of WM capacity.

The key finding in this study was made possible by our use of the whole-report procedure in which all relevant features of each item were reported during each trial. This provided a clear picture of the *distribution* of the stored feature values across the six items in the sample array. We replicated past findings that observers had a strong tendency to report the best remembered items first (Adam et al., 2017; Ngiam et al., 2022; but see Oberauer, 2022) such that accurate recall was concentrated within a subset of three objects with chance performance in the last three responses, indicative of an item limit of approximately 3. Critically, this empirical pattern reveals a strong object-based clustering of all successfully stored feature values to a 3-item subset of the 6-item sample array. This pattern contradicted the *independent feature* model of WM storage, in which separable resources for each feature dimension can be deployed without respect to the objects that contain the stored features (Fougnie & Alvarez, 2011). Accordingly, formal model comparisons indicated our data was best fit by an *object-based pointer* model (asserting object-based storage limits with independent feature loss) for all 120 participants across the four experiments, outperforming the *independent features* model (feature-based memories distributed across the memoranda independent of the objects that contain them). The *object-based pointer* model also provided superior fits compared to the *strong object* model (memory representations are feature-integrated objects without loss), emphasizing that object-based memories are not lossless. Instead, independent feature loss was observed within each working memory representation, in line with past work (Fougnie & Alvarez, 2011). Our modelling results closely align with the findings of Sone et al. (2021) and Li et al. (2022), who reported the strong but imperfect concurrence of recall precision for features from the same item, and also favored an account with integrative object-based representation with independent feature storage (see also Markov et al., 2019). In addition, our results extend evidence for the existence of an object-based item limit in visual WM capacity, with guessing in the last three responses beyond that item limit, replicating the findings of previous whole-report experiments conducted by Adam et al. (2017) and Ngiam et al. (2022) (but see also, Oberauer, 2022).

A viable explanation for the object-based item limit that we observed across all of our experiments is a cap to the number of discretized items that can be stored in WM, possibly because of a limit in the number of *content-independent pointers* that enable tracking of items through time and space (Hakim et al., 2019; Thyer et al., 2022). This object-based *pointer* hypothesis is inspired by the ideas of Pylyshyn's (1989) *fingers of instantiations* (FINSTs) and Kahneman, Treisman and Gibbs' (1992) *object files*. Both theories describe a for tracking individuated items through time and space, with a distinct but parallel mechanism for mapping featural content to the representation of each item. A similar *serial token* mechanism features in the memory for latent representations (MLR) model of Hedayati et al. (2022), whereby specific patterns of activity in a binding pool of neurons are mapped to and retrieved from tokens, enabling individuation of item representations in working memory. This is akin to the explicit conjunctive coding of feature conjunctions that has recently been observed in the human perirhinal cortex (Erez et al., 2016; Liang et al., 2020) and a similar pointer mechanism has been theorized as a way the human brain adapts representations of content to representations of systematic structures (O'Reilly et al., 2022).

Thyer et al. (2022) provided electrophysiological evidence for the existence of content-independent pointers. They reported a multivariate signature in electroencephalography (EEG) activity for WM load that generalizes across changes in both the *type* of visual features stored, and the *number* of features within each to-be-remembered item. Critically, the multivariate load signature respected the number of items rather than the total number of attended locations, suggesting a consistent deployment of this *pointer* mechanism to the objects rather than locations. This dovetails with the findings of Hakim et al. (2019) regarding the contralateral delay activity (CDA), an event-related potential component measured using EEG that has consistently been shown to track working memory load (Balaban & Luria, 2016; Feldmann-Wüstefeld et al., 2018; Ikkai et al., 2010; Luria et al., 2016; Vogel & Machizawa, 2004). Hakim et al. (2019) reported no CDA when spatial attention was deployed to specific locations, but a clear CDA component when items occupying the same number of locations were actively maintained in WM (see also Balaban, Drew and Luria (2019) which similarly reports neural evidence for an object-based pointer account of WM with the CDA). We believe this is giving rise to the observed item limit in our experiments – only three pointers can be reliably deployed at the same time.

An object-based pointer model parsimoniously accommodates for the independent feature loss we observed in our experiments and that also has been documented in the past (Fougnie & Alvarez, 2011). That is, while object-based content is mapped on to the pointers, decay of the mapping from pointers to a feature produces the dropping of independent features from the WM representation and is then unavailable to be retrieved for recall. Cowan et al. (2013) proposed a model very similar in conception to the object-based pointer account proposed here. In their model, WM is assumed to have a fixed object-based capacity limit. Within this limit, while one feature of an object is always remembered, storage of any additional features from the same object is probabilistic and therefore potentially unsuccessful (see Hardman & Cowan, 2015; Oberauer & Eichenberger, 2013 for successful applications). This account would align well with our object-based pointer model if we presume that location is a core (i.e., always stored) feature of each object encoded into working memory (Golomb et al., 2014), and probabilistic loss (or failure to encode) of additional nonspatial features.

A recent study proposed that the mechanism for decay may not necessarily be feature-specific in nature, but the result of accumulating internal noise (Kuuramo et al., 2022). That is to say, feature loss or other *feature-based* effects are produced by factors that impact recall performance for a single item in WM. However, the present study indicates that the overall *storage ceiling* for WM is determined by object-based encoding. Thus, we believe the object-based pointer model to be a productive theoretical framework for understanding capacity limits in working memory (see Ngiam, 2023).

An alternative interpretation of the present results is that the limit on accurate recall is not set by object-based encoding *per se*, but rather by the number of *locations*. By this account, feature values are encoded independently – and without respect to the objects that contain them – but are clustered due to their shared spatial location (Schneegans & Bays, 2017, 2019). On the one hand, this location-based perspective is compatible with the spatiotemporal pointer account that we have proposed, because spatial indexing is a core component of this content-independent indexing operation. On the other hand, one can also conceive of a pure location-based account in which there is no assumption of object-based representations that constrain encoding into WM. In the latter case, both features of a single object might be encoded not because they are contained in the same object, but because those features occupied a shared location. The present results cannot provide a decisive test between these alternatives, but there is past behavioral and neural work that argues against location-based determinants of WM capacity. For instance, CDA activity tracks the total number of objects encoded into visual WM even when multiple objects are sequentially presented in the same position (Ikkai et al., 2010; Vogel et al., 2005). Moreover, visual working memory performance is not affected when multiple stimuli are presented sequentially within the same location, compared to when each stimulus has a unique location (Woodman et al., 2012). Thus, our working hypothesis is that storage in visual WM is constrained by the number of objects stored, rather than by the number of locations those objects occupy. This object-based model of capacity also entails the assumption all features stored from multi-feature objects will also be constrained by the number of objects they occupy rather than the number of positions in which those features appeared.

In sum, our findings provide novel evidence for an object-based limit on storage in visual working memory. Successfully recalled features were restricted to the first three items recalled for both single- and multi-featured objects, contradicting accounts that argue for independent storage of distinct features without regard to the objects that contain those features. In addition, we replicated past findings of object-based benefits, with substantially more feature values recalled from multi-featured objects relative to objects with only one relevant feature. Thus, although our object-based pointer account acknowledges that features can be lost in a probabilistic manner, capacity limits in visual WM appear to be constrained by an *object-based* limit of approximately three items. These findings fall in line with our working hypothesis that storage in visual working memory is constrained by the limited availability of spatiotemporal *pointers* (akin to FINSTs and object files) that enable the tracking of stored items through time and space.

Constraints of Generality

We observed the same empirical pattern across four experiments with differing encoding durations (500 msec in Experiments 1, 3 and 4, and 150 msec in Experiment 2) and stimuli (colored clock faces in Experiments 1 and 2, colored triangles in Experiment 3, and colored shapes in Experiment 4). We note that a factor that can impact performance in visual WM is the similarity between items – we employed highly discriminable colors, orientations and shapes to reduce any effect of similarity. Nevertheless, we expect our key finding – that object-based encoding constrains capacity limits – to generalize across presentation durations and stimuli typically used in visual WM studies.

The four experiments conducted in the present study contained unique samples – no participant had completed multiple experiments. However, the participant samples came from the local University of Chicago community, mostly (but not all) young adults attending college. Our impression is that the demographics of our participants will somewhat match that of the University of Chicago student corpus. While there are individual differences in visual WM, we expect our findings will generalize across samples of the human population.

References

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The Contribution of Attentional Lapses to Individual Differences in Visual Working Memory Capacity. *Journal of Cognitive Neuroscience*, *27*(8), 1601–1616. https://doi.org/10.1162/jocn_a_00811

Adam, K. C. S., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79–97. https://doi.org/10.1016/j.cogpsych.2017.07.001

Balaban, H., Drew, T., & Luria, R. (2019). Neural evidence for an object-based pointer system underlying working memory. *Cortex*, *119*, 362–372. https://doi.org/10.1016/j.cortex.2019.05.008

Balaban, H., & Luria, R. (2016). Integration of Distinct Objects in Visual Working Memory Depends on Strong Objecthood Cues Even for Different-Dimension Conjunctions. *Cerebral Cortex*, *26*, 2093–2104.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 4–4. https://doi.org/10.1167/11.5.4

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

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 731–747. https://doi.org/10.1037/a0029687

Erez, J., Cusack, R., Kendall, W., & Barense, M. D. (2016). Conjunctive Coding of Complex Object Features. *Cerebral Cortex*, *26*(5), 2271–2282. https://doi.org/10.1093/cercor/bhv081

Feldmann-Wüstefeld, T., Vogel, E. K., & Awh, E. (2018). Contralateral Delay Activity Indexes Working Memory Storage, Not the Current Focus of Spatial Attention. *Journal of Cognitive Neuroscience*, 1–11. https://doi.org/10.1162/jocn_a_01271

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12), 1–12. https://doi.org/10.1167/11.12.3

Fougnie, D., Cormiea, S. M., & Alvarez, G. A. (2012). Object-Based Benefits Without Object-Based Representations. *Journal of Experimental Psychology: General*, *142*(3), 621–626. https://doi.org/10.1037/a0030300

Gao, T., Gao, Z., Li, J., Sun, Z., & Shen, M. (2011). The perceptual root of object-based storage: An interactive model of perception and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1803.

Golomb, J. D., Kupitz, C. N., & Thiemann, C. T. (2014). The influence of object location on identity: A “spatial congruency bias.” *Journal of Experimental Psychology: General*, *143*, 2262–2278. https://doi.org/10.1037/xge0000017

Hakim, N., Adam, K. C. S., Gunseli, E., Awh, E., & Vogel, E. K. (2019). Dissecting the Neural Focus of Attention Reveals Distinct Processes for Spatial Attention and Object-Based Storage in Visual Working Memory. *Psychological Science*, *30*(4), 526–540. https://doi.org/10.1177/0956797619830384

Hakim, N., deBettencourt, M. T., Awh, E., & Vogel, E. K. (2020). Attention fluctuations impact ongoing maintenance of information in working memory. *Psychonomic Bulletin & Review*, *27*(6), 1269–1278. https://doi.org/10.3758/s13423-020-01790-z

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*(2), 325–347. https://doi.org/10.1037/xlm0000031

Hedayati, S., O’Donnell, R. E., & Wyble, B. (2022). A model of working memory for latent representations. *Nature Human Behaviour*, *6*(5), Article 5. https://doi.org/10.1038/s41562-021-01264-9

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral Delay Activity Provides a Neural Measure of the Number of Representations in Visual Working Memory. *Journal of Neurophysiology*, *103*(4), 1963–1968. https://doi.org/10.1152/jn.00978.2009

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kuuramo, C., Saarinen, J., & Kurki, I. (2022). Forgetting in visual working memory: Internal noise explains decay of feature representations. *Journal of Vision*, *22*(8), 8. https://doi.org/10.1167/jov.22.8.8

Li, A. Y., Fukuda, K., & Barense, M. D. (2022). Independent features form integrated objects: Using a novel shape-color “conjunction task” to reconstruct memory resolution for multiple object features simultaneously. *Cognition*, *223*, 105024. https://doi.org/10.1016/j.cognition.2022.105024

Liang, J. C., Erez, J., Zhang, F., Cusack, R., & Barense, M. D. (2020). Experience Transforms Conjunctive Object Representations: Neural Evidence for Unitization After Visual Expertise. *Cerebral Cortex*, *30*(5), 2721–2739. https://doi.org/10.1093/cercor/bhz250

Lin, Y., Kong, G., & Fougnie, D. (2021). Object-based selection in visual working memory. *Psychonomic Bulletin & Review*, *28*(6), 1961–1971.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews*, *62*, 100–108. https://doi.org/10.1016/j.neubiorev.2016.01.003

Markov, Y. A., Tiurina, N. A., & Utochkin, I. S. (2019). Different features are stored independently in visual working memory but mediated by object-based representations. *Acta Psychologica*, *197*, 52–63.

Ngiam, W. X. Q. (2023). *Mapping visual working memory models to a theoretical framework*. PsyArXiv. https://doi.org/10.31234/osf.io/g8erx

Ngiam, W. X. Q., Foster, J. J., Adam, K. C. S., & Awh, E. (2022). Distinguishing guesses from fuzzy memories: Further evidence for item limits in visual working memory. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-022-02631-y

Oberauer, K. (2022). Little Support for Discrete Item Limits in Visual Working Memory. *Psychological Science*, *33*(7), 1128–1142. https://doi.org/10.1177/09567976211068045

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition*, *41*(8), 1212–1227. https://doi.org/10.3758/s13421-013-0333-6

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, *64*(7), 1055–1067. https://doi.org/10.3758/BF03194756

O’Reilly, R. C., Ranganath, C., & Russin, J. L. (2022). The Structure of Systematicity in the Brain. *Current Directions in Psychological Science*, *31*(2), 124–130. https://doi.org/10.1177/09637214211049233

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *The Journal of Neuroscience*, *37*, 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schneegans, S., & Bays, P. M. (2019). New perspectives on binding in visual working memory. *British Journal of Psychology (London, England: 1953)*, *110*(2), 207–244. https://doi.org/10.1111/bjop.12345

Schor, D., Brodersen, A. S., & Gibson, B. S. (2020). A model comparison approach reveals individual variation in the scope and control of attention. *Psychonomic Bulletin & Review*, *27*(5), 1006–1013. https://doi.org/10.3758/s13423-020-01744-5

Schor, D., Wilcox, K. T., & Gibson, B. S. (2023). Partial recall: Implications for the discrete slot limit of working memory capacity. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-023-02713-5

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Shin, H., & Ma, W. J. (2017). Visual short-term memory for oriented, colored objects. *Journal of Vision*, *17*(9), 12–12. https://doi.org/10.1167/17.9.12

Sone, H., Kang, M.-S., Li, A. Y., Tsubomi, H., & Fukuda, K. (2021). Simultaneous estimation procedure reveals the object-based, but not space-based, dependence of visual working memory representations. *Cognition*, *209*, 104579. https://doi.org/10.1016/j.cognition.2020.104579

Thyer, W., Adam, K. C. S., Diaz, G. K., Velázquez Sánchez, I. N., Vogel, E. K., & Awh, E. (2022). Storage in Visual Working Memory Recruits a Content-Independent Pointer System. *Psychological Science*, *33*(10), 1680–1694. https://doi.org/10.1177/09567976221090923

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. https://doi.org/10.1038/nature02447

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

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*(1), 92.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64. https://doi.org/10.1037/0096-3445.131.1.48

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2012). Flexibility in Visual Working Memory: Accurate Change Detection in the Face of Irrelevant Variations in Position. *Visual Cognition*, *20*(1), 1–28. https://doi.org/10.1080/13506285.2011.630694