

Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model

Daryl Fougnie

Department of Psychology, Harvard University,
Cambridge, MA, USA



George A. Alvarez

Department of Psychology, Harvard University,
Cambridge, MA, USA



The world is composed of features and objects and this structure may influence what is stored in working memory. It is widely believed that the content of memory is object-based: Memory stores integrated objects, not independent features. We asked participants to report the color and orientation of an object and found that memory errors were largely independent: Even when one of the object's features was entirely forgotten, the other feature was often reported. This finding contradicts object-based models and challenges fundamental assumptions about the organization of information in working memory. We propose an alternative framework involving independent self-sustaining representations that may fail probabilistically and independently for each feature. This account predicts that the degree of independence in feature storage is determined by the degree of overlap in neural coding during perception. Consistent with this prediction, we found that errors for jointly encoded dimensions were less independent than errors for independently encoded dimensions.

Keywords: memory, attention, visual cognition

Citation: 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):3, 1–12, <http://www.journalofvision.org/content/11/12/3>, doi:10.1167/11.12.3.

Introduction

We process the visual world not as a collection of features but as a set of meaningful objects composed of those features. The object-based structure of visual cognition is evident in attention (Scholl, 2001)—e.g., featural information follows the spatiotemporal properties of objects (Kahneman, Treisman, & Gibbs, 1992; Mitroff & Alvarez, 2007) and attention spreads not in empty space but in space as defined by objects (Egley, Driver, & Rafal, 1994). Visual working memory is also influenced by the organization of features into objects. For example, working memory performance for multiple features depends on whether features are grouped into objects (Luck & Vogel, 1997; Xu, 2002). These findings have led to the view that we encode and maintain integrated object representations (Cowan, 2001; Luck & Vogel, 1997; Rensink, 2000).

Here, we test such object-based models by determining how well performance for one feature is correlated with memory for other features of the same object. A pure object-based model predicts that memory for different task-relevant features of the same object will be highly correlated, and if one feature is unknown, the other features will also be unknown. In contrast, a pure feature-based model predicts that memory for one feature will be completely independent of memory for another feature of the same object. We found that the color and orientation reports for the same object were largely

independent. Participants often knew the color of an object that they did not know the orientation of and *vice versa*. These results suggest that the contents of working memory are not integrated object representations. We propose that features that are coded independently during perception will fail probabilistically and independently in visual working memory.

Experiment 1

Methods

Participants

Twelve volunteers (18–35 years old) participated for \$10/h or course credit.

Stimuli

Each display showed five isosceles triangles spaced equally in a ring, 2.5° (visual angle) in radius, around a central fixation (Figure 1A). Each triangle had angles of 30° , 75° , and 75° and sides subtending $0.7^\circ \times 1.63^\circ \times 1.63^\circ$ (visual angle). Each triangle had a randomly chosen orientation (2° – 360° , in 2° steps) and color (from 180 equiluminant colors evenly distributed along a circle in the CIE $L^*a^*b^*$ color space centered at $L = 54$, $a = 18$, $b = -8$, with a radius of 59).

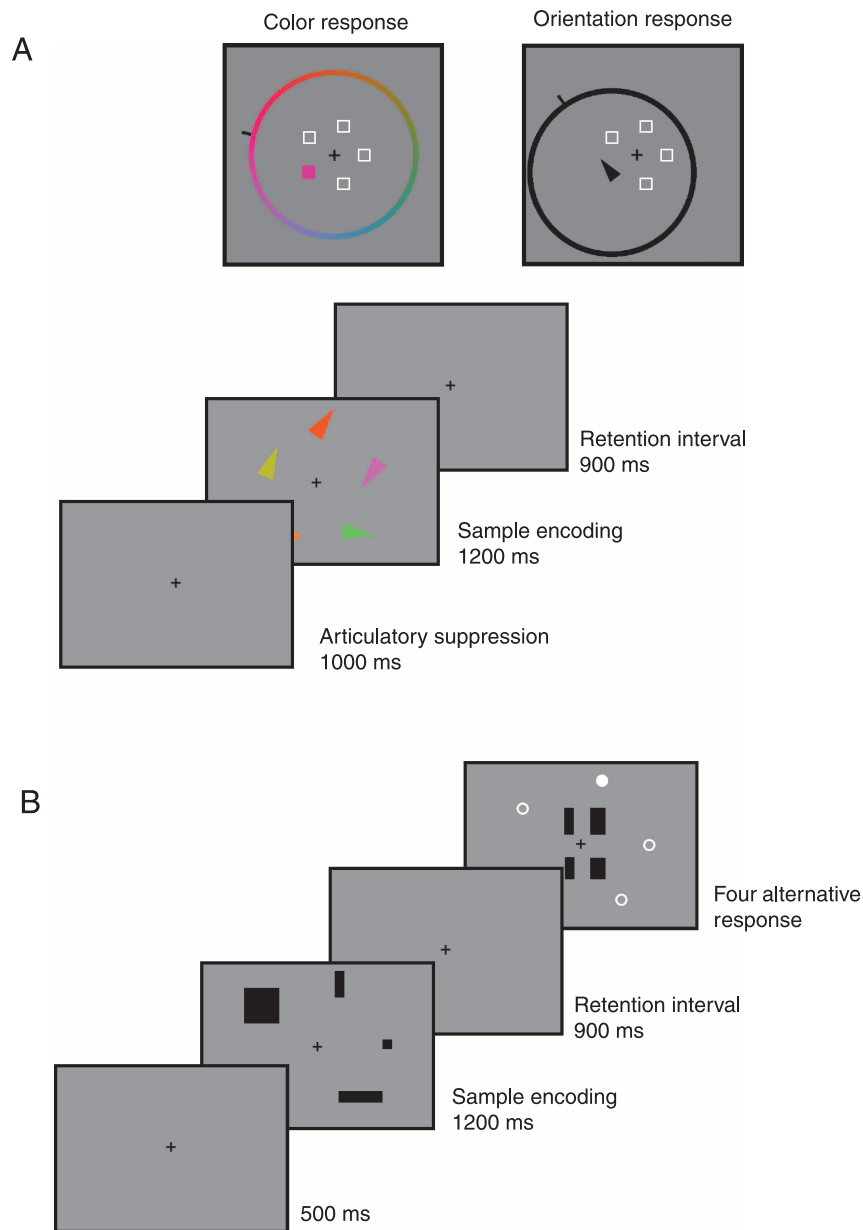


Figure 1. (A) Trial timeline for Experiment 1. (B) Trial timeline for Experiment 2A.

Procedure

Trials began with the presentation of a central cross (1 s), followed by the memory display (1200 ms). After a retention interval (900 ms), a filled white squares appeared at the probed triangle's location, and hollow squares appeared at non-probed locations. Participants were asked to report the color and then the orientation of the probed item or *vice versa* (randomly determined). For color reports, a response wheel of 180 color segments appeared centered around the probe display. For orientation reports, a black response wheel was centered on the probed item (Figure 1A). Participants selected one of 180 values for each report by clicking the mouse. An indicator line outside the response wheels, with position determined

by cursor position, indicated the selected value. Once the participants moved the mouse, the probed location contained a colored square (color reports) or black triangle (orientation reports) with the feature matching the selected value. Responses were unsped and error feedback was given. Each participant completed 540 trials. Six participants were also monitored for articulatory suppression to minimize verbal encoding and rehearsal by repeating the word *the* three times per second for each trial duration.

Analysis

A modeling analysis was first performed with all trials to classify guess trials, providing measures of color

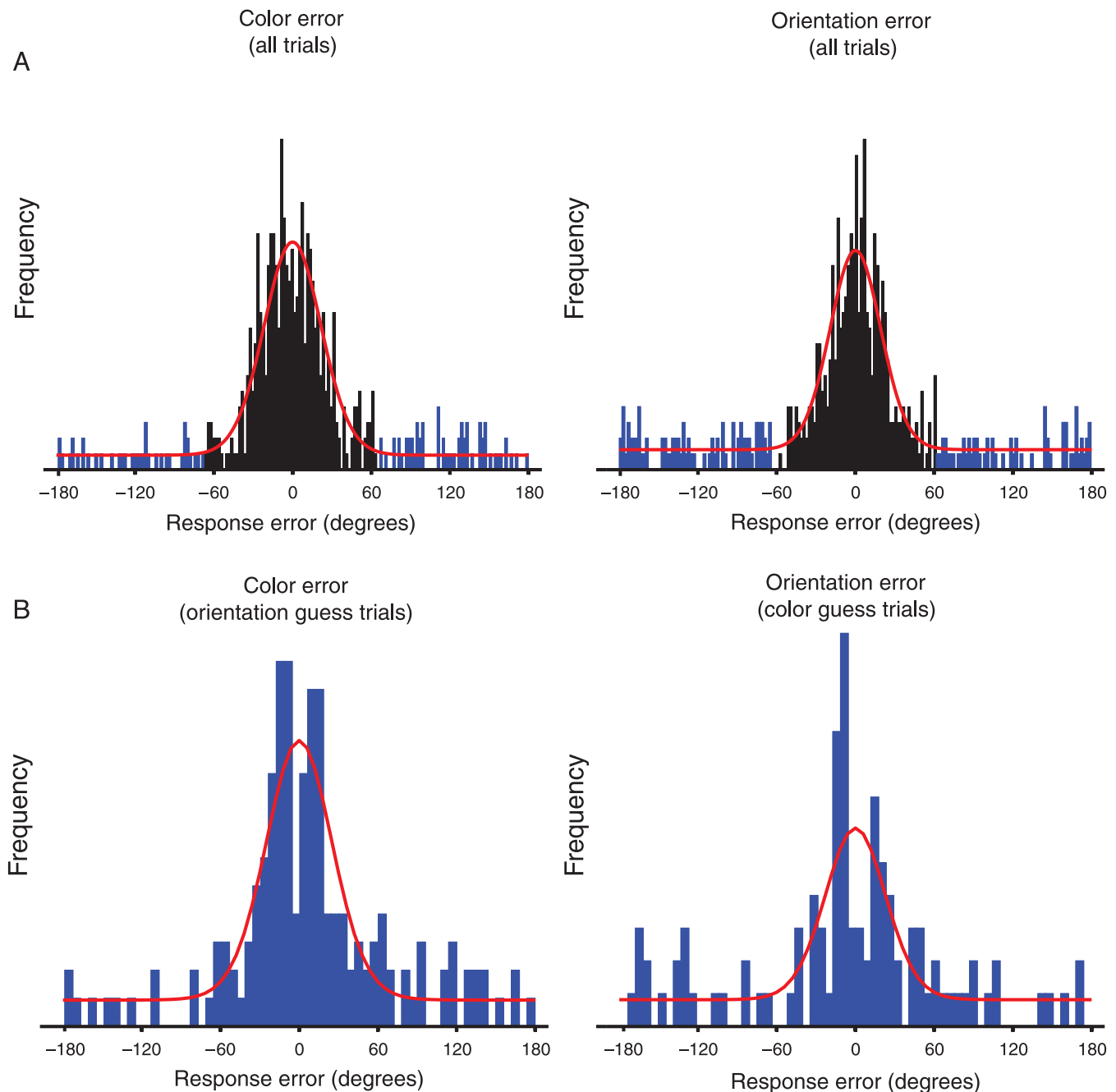


Figure 2. (A) Response error histograms for color (left) and orientation (right) reports for a single observer. The blue shaded portion of the histogram indicates responses at least three standard deviations away from the target (guess trials). (B) Guess trials were used to generate response error histograms, for the other feature, for color (left) and orientation (right) responses (filter condition). The red line indicates the best fitting mixture model of each condition.

memory when observers guessed about orientation and *vice versa*. Response error was calculated by subtracting each probed item's correct value from the response. The response error distributions were fit using maximum likelihood estimation as a mixture of a circular normal distribution centered on the target value, a uniform distribution, and a mixture of circular normal distributions for each of the four non-probed items centered on that item's value. The height parameters of these distributions

represented estimates of the percentage of target responses, random guesses, and swap errors, respectively. In addition, the model had a standard deviation parameter for the width of the normal distributions that provided an estimate of the precision of stored representations.

To examine whether memory failures were independent for orientation and color, mixture modeling was also performed for the filtered response distributions by including the subset of trials in which participants

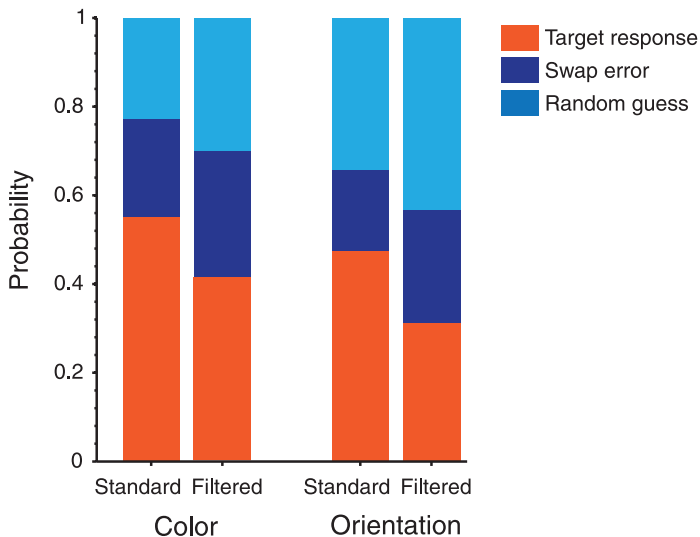


Figure 3. Model estimates of the proportion of target responses (orange), swap errors (blue), and random guesses (light blue) for the standard and filtered conditions for both color (left) and orientation (right) responses.

responded more than 3 standard deviations away from the target value of the other feature (Figure 2A). This criterion was selected because target responses are unlikely to fall outside this range (0.26%), while this criterion still afforded a sufficient number of trials to perform the mixture modeling analysis. Similar results were observed when responses were filtered by 2 or 4 standard deviations.

Results and discussion

Contrary to the object-based prediction, the filtered response distributions were not uniform but showed a large proportion of responses distributed around the correct value (Figure 2B), suggesting independent failures of feature memory. A comparison of the parameter fits (Figure 3) found that the subset of trials where participants guessed on the other feature (filtered trials) had less target responses and more swap errors and random guesses ($ps < 0.05$; see Supplementary data for ANOVA results for each model parameter). Strikingly, while performance was slightly worse for the subset of filtered trials, the difference was much less than the all-or-none object storage predicted by pure object-based models of memory. To quantify memory independence between features, we developed a measure termed the separability index (SI) that takes a ratio of the proportion of target responses in the filtered trials to the proportion of target responses in all trials (standard condition, averaged across features). The higher the separability index, the greater the independence of memory for the features. Pure object-based models of working memory predict an SI of 0, whereas

pure feature-based models predict an SI of 1, as the organization of features into objects has no effect on memory. The SI value (0.73 ± 0.04) was much closer to the prediction of the feature-based model than of the object-based model (Figure 4), showing largely independent working memory for the two features. Similarly, a scatter plot of color and orientation errors reveals largely independent performance across features (Figure 5). In fact, the absolute magnitude of error for the two responses was only weakly correlated (the average of r^2 values was 0.03).

These results are incompatible with theories suggesting that we encode and maintain a subset of integrated objects (Cowan, 2001; Luck & Vogel, 1997). Moreover, these results cannot be explained by failures of feature binding in working memory (Brown & Brockmole, 2010; Fougnie & Marois, 2009; Wheeler & Treisman, 2002), where features from non-probed items may be mistakenly reported instead of the probed item. This is revealed by an additional analysis that removed responses that were within 2 SDs of any sample item's value and still showed a high proportion of target responses (Supplementary Figure 1; SI value of 0.67) contrary to the misbinding account.

Further control analyses showed that the independent failures were not due to verbal rehearsal or response interference. The participants who performed an articulatory suppression task had an SI of 0.74 ± 0.05 , which did not differ from the SI of 0.70 ± 0.06 for the other participants ($p = 0.41$). In addition, SI values did not differ if filtered for the first (0.76 ± 0.03) versus the second (0.71 ± 0.04) response showing that feature independence is not due to response order effects.¹ The SI values were not significantly altered if the canonical orientations (e.g., upright) and colors (e.g., green) are excluded (see Supplementary data). We also replicated the results of independent failures in another experiment ($n = 8$) using a task that was identical to Experiment 1, except that participants adjusted both the

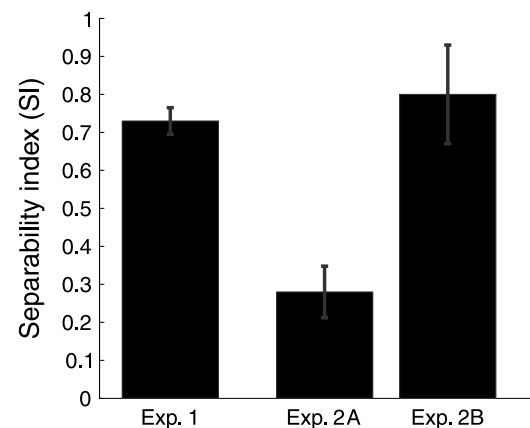


Figure 4. Ratio of proportion of target responses in the filtered condition to the standard condition (separability index) for Experiments 1 and 2.

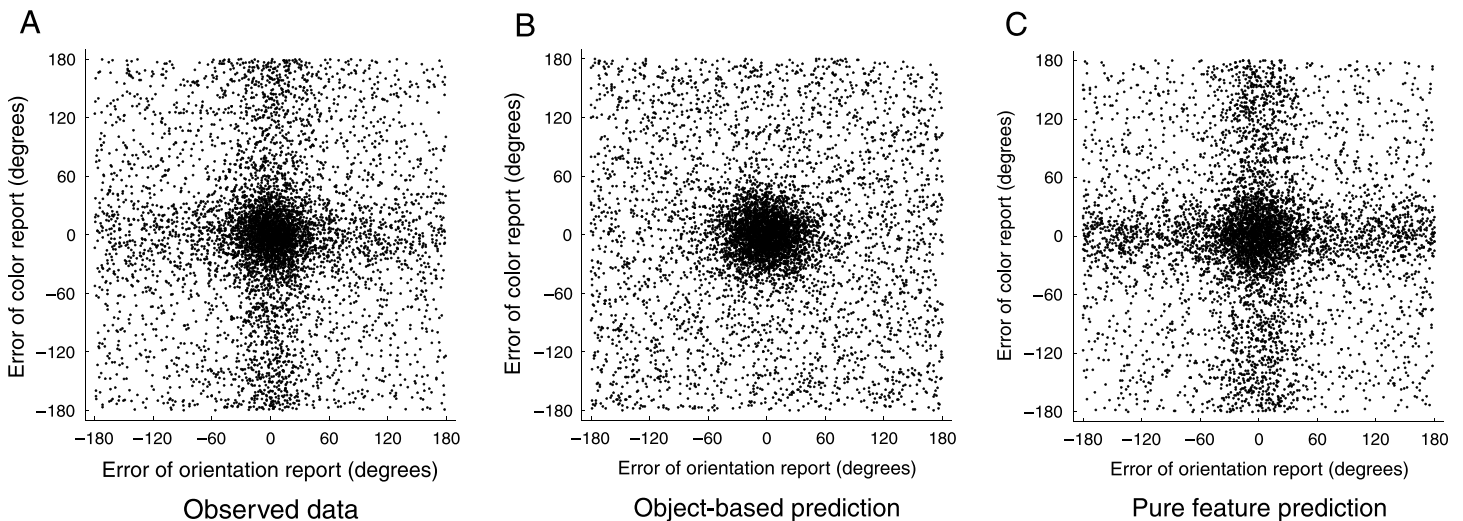


Figure 5. (A) Scatter plot of all color and orientation responses in [Experiment 1](#). (B) Simulated scatter plot data from object-based model responses are either correlated with the target value for both features or randomly distributed for both features. (C) Simulated scatter plot data from a feature-based model—the likelihood of each response being correlated with the target or randomly distributed is independent for each feature.

color and orientation of an isosceles triangle to make a single response (SI of 0.90 ± 0.07).

Here, participants often failed to remember color or orientation, as shown by the large proportion of random guesses. Moreover, errors for these features were largely independent for a single object. How do we reconcile this with the extensive literature showing that working memory is sensitive to the object-based structure of a display? (Delvenne & Bruyer, 2004, 2006; Luck & Vogel, 1997; Olson & Jiang, 2002; Xu, 2002). We suggest that these random guesses may arise from probabilistic failures of self-sustaining neural networks (Amit, Brunel, & Tsodyks, 1994; Hebb, 1949; Johnson, Spencer, Luck, & Schoner, 2008; Johnson, Spencer, & Schöner, 2009; Rolls & Deco, 2010). In this framework, working memory is the maintenance of perceptual representations in the absence of bottom-up sensory input. We suggest that increased item or information load may lead to detrimental effects on working memory due to an increased likelihood of representational failure. However, the representational unit of memory failure is not at the level of the object, at least for objects defined by color and orientation. Consider that successful visual working memory is the sustained activation of representations in the absence of bottom-up perceptual input. Since orientation and color are encoded by largely independent neurons during perception (i.e., separable feature dimensions; Cant, Large, McCall, & Goodale, 2008; Garner, 1974), these features may form largely independent self-sustaining representations and these representations may fail independently.² Critically, this *probabilistic feature-store* account proposes that the degree of independence for features in working memory is determined by the degree of overlap

in neural coding of the features. If a task requires maintenance of two features that are coded by overlapping populations of neurons, then their representations will overlap in memory and will not fail independently. This prediction was assessed in [Experiment 2](#).

Experiment 2

Unlike color and orientation, height and width are considered integral feature dimensions (Cant et al., 2008; Dykes & Cooper, 1978; Ganel & Goodale, 2003) and overlap in neural coding. The probabilistic feature-store model proposed here predicts that memory for jointly coded features (height/width, Experiment 2A) will be more correlated than memory for independently coded features (color/orientation, Experiment 2B).

Unlike [Experiment 1](#), which modeled precision and guess rate via a continuous report task, [Experiment 2](#) modeled these parameters using performance in a recognition task, which is useful when the feature dimensions are not circular like color and orientation (Bays & Husain, 2008). The slope of the function relating performance with the difference between target and non-target probes reflects the precision of memory (more precision, steeper slopes). The level of performance at the largest probe difference reflects the percentage of guess and non-guess responses. Thus, while [Experiments 1](#) and [2](#) differ in terms of task, the underlying model (where participants either guess or respond with imprecise target knowledge) is identical.

Methods

Participants

Sixteen volunteers (18–35 years old; 8 each in Experiments 2A and 2B) participated for \$10/h or course credit.

Stimuli

Each memory display in Experiment 2A showed five rectangles with heights and widths selected randomly from 180 evenly spaced values between 0.5° and 10° presented in equally spaced positions along an imaginary ring 7° from fixation. Experiment 2B had identical stimulus parameters as in Experiment 1 except that the display radius was increased to 7° as in Experiment 2A.

Procedure

Participants selected the item corresponding to the probed location from four alternatives: the correct response, two foils that matched the probe in one feature, and a foil matching neither feature (Figure 1B). The foil values were independent and differed from the target by $0.4^\circ/1.5^\circ/3.5^\circ/5^\circ$ of visual angle in Experiment 2A and $180^\circ/140^\circ/30^\circ/10^\circ$ of feature space in Experiment 2B. For example, if the color and orientation choices differed by 180° and 30° , respectively, then the four response alternatives were the full cross of the correct and incorrect color and orientation choices such that each feature judgment was completely orthogonal. Participants selected among the four options using response keys (“E”/“D”/“I”/“K”), which corresponded with the onscreen positions of the response options. Each participant completed 512 trials. Error feedback was given separately for each feature.

Analysis

Percent correct for each feature judgment (proportion of trials that participants selected the item with the correct value in that feature) was measured as a function of the difference between the correct and incorrect feature values. Intuitively, if items are stored with high precision, then it would be easy to select the correct choice even with a small difference, and the function would be very steep. If items were stored with low precision, then large changes would be necessary to correctly respond, and the function would be shallow. Thus, the steepness of the curve reflects the precision of memory. On the other hand, if the observer has no information and is guessing, then even the largest possible change would not have perfect performance. For example, if observers remembered half of the items performance would asymptote at 75% correct (getting it correct half the time, and randomly guessing half the time). To quantify these parameters (precision and guessing), performance function was fit to a mixture of a

uniform distribution and a cumulative Gaussian, with the mixture parameter providing an estimate of the proportion of target responses ($1 -$ the proportion of guesses) and the Gaussian standard deviation providing an estimate of memory precision (Zhang & Luck, 2008). Monte Carlo simulations found that this method provided better parameter estimates with low numbers of trials than other modeling approaches for recognition data (e.g., Bays & Husain, 2008).

Since discrimination tasks produce less reliable parameter estimates than report tasks, we used a more conservative criterion of guess trials in this experiment than Experiment 1. We assumed that observers guessed whenever they incorrectly judged a feature when the target–foil deviations were at least 3.5° in visual angle for Experiment 2A or were at least 140° in feature space (70 color steps) for Experiment 2B. A potential concern for Experiment 2A is that the large target–foil deviations may still not have been sufficiently large to be outside of the precision of stored representations. Thus, some of the filtered trials may have been target response trials. Note, however, that this works against our hypothesis and findings—if observers were not guessing on height or width, there is no reason to suspect that they would have no information about the other feature.

Results and discussion

Participants’ performance for height and width judgments (Figure 6A) for all trials and for the subset of trials where participants guessed on the other feature (filtered trials) were modeled as a mixture of a cumulative Gaussian distribution and a uniform distribution to determine the proportion of target response trials (see Methods section). We found that participants rarely had information about height or width without information about the item’s other feature, as the proportion of target responses for the filtered trials (12%) was significantly less than for all trials (37%; $t(7) = 5.24, p < 0.005$).

In contrast, we replicated the finding of largely independent color and orientation memory in Experiment 2B using the same task as in Experiment 2A (Figure 6B). The proportion of target responses after filtering (26%) did not significantly differ from the standard condition (33%; $t(7) = 1.63, p = 0.15$) but may have with a larger sample size. Importantly, comparisons between Experiments 2A and 2B show that independent failures of memory were more common for color and orientation working memory. The SI value for Experiment 2A (0.28 ± 0.07) was much smaller than that for Experiment 2B (0.80 ± 0.13 ; $t(14) = 4.09, p < 0.001$). Comparisons of correct judgments between the standard and filtered trials also support a difference between studies. For height and width judgments, overall performance was 59.6% correct, but it dropped significantly ($p < 0.05$) after filtering (52.5%, no longer above chance).

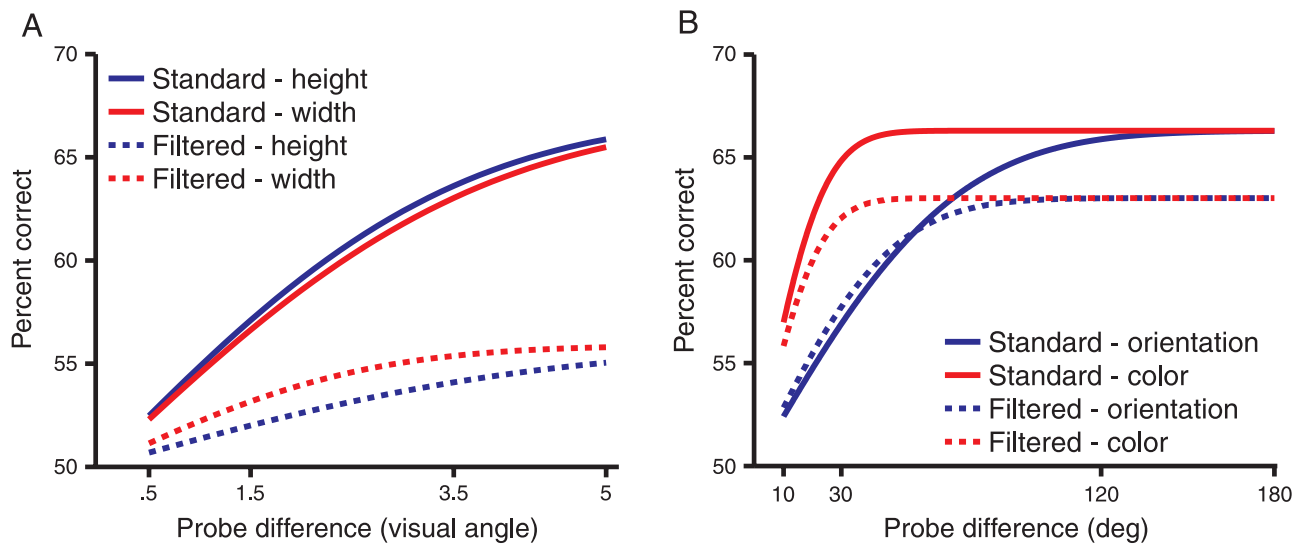


Figure 6. Model estimates of task performance as a function of the difference between target and foil values for (A) height and width (Experiment 2A) or (B) color and orientation (Experiment 2B; parameters were averaged from the best fit for each participant). The y-axis represents the percentage of trials in which participants chose a probe item matching in the relevant feature. Increasing deviations between target and foil are plotted from left to right on the x-axis.

For color and orientation, performance in the standard and filtered trials did not differ significantly (59.6% vs. 58.4%; $t(7) = 0.83$, $p = 0.43$). Indeed, performance costs after filtering were significantly greater for height and width, as confirmed by a between-subjects ANOVA with factors of Condition (standard, filtered) and Feature pairing (color–orientation, height–width; $F(1, 14) = 9.25$, $p < 0.01$). Thus, both model fits and model-independent estimates of performance converge in showing that working memory for color and orientation is largely independent, while working memory for integral features like height and width is not.

General discussion

The theory that integrated objects are the primary representational unit in visual cognition and working memory has been highly influential in psychology, neuroscience, and computational modeling (Cowan, 1988, 2001; Hollingworth & Rasmussen, 2010; Johnson et al., 2008; Kahneman et al., 1992; Luck & Vogel, 1997; Mitroff & Alvarez, 2007; Rensink, 2000; Scholl, 2001; Vogel, Woodman, & Luck, 2001; Wolters & Raffone, 2008; Zhang & Luck, 2008). The present findings challenge the assumption that the representational unit of working memory is always an integrated object (Lee & Chun, 2001; Luck & Vogel, 1997; Rensink, 2000; Vogel et al., 2001; Wolters & Raffone, 2008; Zhang & Luck, 2008). When the task was to remember the color and orientation of items, we found that participants were quite

accurate at indicating the orientation of an object even when they had guessed the color of the *same* object and vice versa. In contrast, we found less independence between responses for the height and width of remembered items. This pattern of results is consistent with previous research on integral and separable perceptual dimensions. Specifically, color and orientation are processed by separate neural populations (Hubel & Wiesel, 1968; Livingstone & Hubel, 1988), and our perception of orientation is largely independent of our perception of color (Garner, 1974). In contrast, height and width draw on overlapping neural populations and cannot be processed independently (Drucker, Kerr, & Aguirre, 2009; Ganel et al., 2006; Garner, 1974). It appears that objects in working memory are collections of features—some integral and some separable—and that the separable feature dimensions are more likely to fail independently than integral feature dimensions.

A probabilistic feature-store model of visual working memory

In this study, participants often responded with no information about the probed target. Indeed, such random guesses have been a common observation in similar tasks (Bays, Catalao, & Husain, 2009; Fougnie, Asplund, & Marois, 2010; Zhang & Luck, 2008). What is the source of such memory failures? Object-based theories suggest that failures occur because only a finite number of items can be stored at a given time—in other words, that there is a structural limit on memory (Cowan, 2001; Luck & Vogel, 1997; Zhang & Luck, 2008). However, working

memory errors may also arise from probabilistic failures. Consider the working memory task in Figure 1A. During perception of the sample display, the featural information of each object is kept stable by a continuous bottom-up signal. However, once the sample is removed, these representations need to self-sustain in the absence of perceptual input. We propose, consistent with computational implementations of self-sustaining neural networks (Amit et al., 1994; Tegner, Compte, & Wang, 2002; Wang, 2001), that these representations are volatile and may fail probabilistically. A key feature of the probabilistic model is that the likelihood of a representation failing is independent of the state of other representations. Furthermore, since orientation and color are processed by distinct neural populations, we propose that the memory of these features may require the self-sustained activation of distinct regions of sensory cortex and that these networks may fail independently of each other.

An alternative account of independent feature memory is provided by the independent feature-store model (Wheeler & Treisman, 2002) in which distinct feature stores each have their own limited capacity. On this account, independent failures could arise due to selectively encoding features from different objects. Although consistent with independent feature storage, selective encoding of object features cannot explain object-based benefits showing that it is easier to remember features grouped into objects (Olson & Jiang, 2002; Xu, 2002). However, both independent feature failures and object-based benefits can be explained by a probabilistic feature-store model in which the likelihood of probabilistic failures increases as the number of objects stored increases. Indeed, the likelihood of feature failures may be differentially affected by increased feature and object load (Fougnie et al., 2010). This probabilistic feature-store model is biologically plausible. For instance, a feature representation may consist of a self-sustaining population of neurons (Ma, Beck, Latham, & Pouget, 2006) and increased feature load may reduce the number of neurons for each population. Increased information load could lead to increased representational failure if the likelihood of representational failure is inversely proportional to the number of neurons that can be devoted to maintaining each unit of information (Tegner et al., 2002). An increased number of distinct objects could also lead to increased representational volatility due to the cost of keeping representations encapsulated such that information from multiple representations will not mutually interfere. This account is admittedly speculative, and future work on this issue is necessary. The important point for present purposes is that object-based encoding benefits are consistent with a probabilistic feature-store model in which maintaining additional objects comes at a cost. Our results suggest that, even though the number of objects may affect the likelihood of failure, the level at which memory representations fail is often the individual feature.

This probabilistic feature-store account not only predicts that failure to remember one feature of an item may be a poor predictor for the memory of the item's other features but also predicts that failure to remember an item from a display will be a poor predictor of performance for the other items. In contrast, if random guesses arose due to limits in the number of representations or features, or because some representations were not given any resources, then working memory performance for one item would be negatively correlated with performance for the other items. For example, if an observer can only store three items, then storage of one item implies that the remaining items are competing for only two spots. Consistent with the probabilistic feature-store account, a recent study has shown that working memory performance is independent across items in a trial (Huang, 2010). We have also replicated this finding using a color and orientation working memory task where participants report the color and orientation of two *different* items (SI value of 1.01). These findings argue against structural limitations as sufficient for explaining failures of working memory.

Capacity limitations versus forgetting

Here, we examine failures of memory for distinct features of the same object. We should be clear that we cannot differentiate whether these failures arise during consolidation into memory or during forgetting from memory. One possible interpretation of these findings is to suggest that studies showing object-based benefits for working memory reflect limitations in how much information can be encoded but that the current evidence of independence arises from additional feature-based forgetting. However, there are several points that argue against this hypothesis. Working memory is much too stable over time to explain the strength of the observed independence between feature reports (Laming & Laming, 1992; Lee & Harris, 1996; Magnussen & Greenlee, 1999; Zhang & Luck, 2009). Indeed, only a modest drop in memory performance was observed in a color report task using a retention interval an order of magnitude larger than the current study (Zhang & Luck, 2009). In addition, we should emphasize that the present methodology is similar to past studies that have found object-based benefits in working memory (Fougnie et al., 2010; Luck & Vogel, 1997; Xu, 2002). Therefore, those findings likely also contained independent failures of object features. If this were the case, the standard interpretation of object-based effects, that performance reflects a number of stored integrated object representations, would be incorrect. In our study, given the degree of independence and the rate of guessing, participants were storing at least one feature from nearly all five items. Thus, even if there is an upper structural limit in the total number of objects that can be stored (Anderson, Vogel, & Awh, 2011; Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997; Zhang & Luck, 2008;

but see Alvarez & Cavanagh, 2004; Wilken & Ma, 2004), this number may have been underestimated in past studies due to the fact that those studies did not account for probabilistic, independent representational failures.

Relationship to previous research on the units of working memory

Studies differ on whether an encoding duration of about 50 ms per item is sufficient (Vogel, Woodman, & Luck, 2006) or whether longer durations (Bays et al., 2009; Eng, Chen, & Jiang, 2005), like those used in Experiments 1 and 2, are necessary for full encoding of the sample display. We replicated feature independence in memory using a 300-ms encoding duration (SI of 0.68) showing that feature independence in memory is observed over the range of expected encoding durations and is not simply an artifact of long encoding durations. However, if participants are unable to encode all items, this will appear as object-based representations (participants will guess for both features of non-encoded objects). This principle might explain why a previous study reported integrated color–shape memory representations (Gajewski & Brockmole, 2006). That study required participants to encode six objects in 187 ms. Furthermore, the display radius was large and may have required shifts of attention to resolve crowding (Bouma, 1970). Indeed, using similar display parameters, we found that 187 ms was not a sufficient encoding duration (change detection performance improved with increased encoding duration, $p < 0.05$). Furthermore, we observed featural independence with this display for sequential (SI of 0.69) but not simultaneous presentations (SI of 0.31), consistent with the presence of encoding limitations during simultaneous presentation.

We found that failures of memory for color and orientation were largely but not completely independent. The lack of complete independence has several potential explanations. One possibility is that representations may sometimes be lost at the object level in addition to the feature level. However, partial independence could be explained by partial overlap in neural coding for color and orientation or by variation in the quality in which items are encoded. For example, the last attended item may be encoded with greater fidelity (Hollingworth & Henderson, 2002). The features of this item would still fail independently, but since they would be less likely to fail than features of other objects, failures would not be completely independent.

Independence between the orientation and color of memory representations has also been reported in a recent study (Bays, Wu, & Husain, 2011). The authors suggested that this independence occurred because participants often reported the color and location of the non-probed item. Against this account, we show that independence occurred even for the subset of trials where participants responded

far away from any item in the sample array (in such cases, misreporting the wrong item is not a plausible explanation for an inaccurate response). Furthermore, we show that this independence depends on the degree to which features are coded independently during perception.

Conclusion

A fundamental assumption of object-based accounts of working memory is that representations are all-or-none and that the failure to remember one feature of an object is necessarily accompanied by the failure to remember other task-relevant features of the object. Here, we have shown that this assumption is incorrect: Representations can independently fail at the feature level. To account for this, we propose that object representations in working memory are composed of a collection of integral and separable feature dimensions and that feature representations fail probabilistically. We also propose that the proximate cause of working memory failures arises from probabilistic failures of self-sustaining representations in the cortical regions that were involved in the initial perception of the objects. This explanation is biologically plausible and draws inspiration from recent neuroimaging findings in perceptual memory (Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009) and from computational models of self-sustaining networks (Amit et al., 1994; Hebb, 1949).

Acknowledgments

We thank Christian Fohrby and Sarah Cormiea for help with data collection. We are grateful to Jordan Suchow, Justin Junge, Ed Awh, and Liqiang Huang for comments on an earlier draft of the manuscript. This work was supported by NIH Grant 1F32EY020706 to D.F. and NIH Grant R03-086743 to G.A.A.

Commercial relationships: none.

Corresponding author: Daryl Fougnie.

Email: darylfougnie@gmail.com.

Address: Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138, USA.

Footnotes

¹For color, non-guess rate was significantly higher for the first (61%) than the second (50%) response, $p < 0.005$. For orientation, non-guess rate was higher for the first

(51%) than the second (46%) response, but this difference was not significant, $p = 0.14$.

²Note that the term failure here is *not* meant to imply that memory representations are lost due to time-based forgetting in working memory. Failure of representations could arise also if there was a failure to consolidate information before the perceptual trace diminished.

References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106–111. [PubMed]
- Amit, D. J., Brunel, N., & Tsodyks, M. V. (1994). Correlations of cortical Hebbian reverberations: Theory versus experiment. *Journal of Neuroscience, 14*, 6435–6445. [PubMed]
- Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *Journal of Neuroscience, 31*, 1128–1138. [PubMed]
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*, 622–628. [PubMed]
- Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10):7, 1–11, <http://www.journalofvision.org/content/9/10/7>, doi:10.1167/9.10.7. [PubMed] [Article]
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851–854. [PubMed]
- Bays, P. M., & Wu, E. Y., Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia, 49*, 1622–1631.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178. [PubMed]
- Brown, L. A., & Brockmole, J. R. (2010). The role of attention in binding visual features in working memory: Evidence from cognitive ageing. *Quarterly Journal of Experimental Psychology, 63*, 2067–2079.
- Cant, J. S., Large, M. E., McCall, L., & Goodale, M. A. (2008). Independent processing of form, colour, and texture in object perception. *Perception, 37*, 57–78. [PubMed]
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*, 163–191. [PubMed]
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–114; discussion 114–185. [PubMed]
- Delvenne, J. F., & Bruyer, R. (2004). Does visual short-term memory store bound features? *Visual Cognition, 11*, 1–27.
- Delvenne, J. F., & Bruyer, R. (2006). A configural effect in visual short-term memory for features from different parts of an object. *Quarterly Journal of Experimental Psychology, 59*, 1567–1580. [PubMed]
- Drucker, D. M., Kerr, W. T., & Aguirre, G. K. (2009). Distinguishing conjoint and independent neural tuning for stimulus features with fMRI adaptation. *Journal of Neurophysiology, 101*, 3310–3324. [PubMed]
- Dykes, J. R., & Cooper, R. G. (1978). An investigation of the perceptual basis of redundancy gain and orthogonal interference for integral dimensions. *Perception & Psychophysics, 23*, 36–42. [PubMed]
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review, 12*, 1127–1133. [PubMed]
- Egley, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*, 161–177. [PubMed]
- Fougnie, D., & Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Visual Cognition, 17*, 48–66. [PubMed]
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision, 10*(12):27, <http://www.journalofvision.org/content/10/12/27>, doi:10.1167/10.12.27. [PubMed] [Article]
- Gajewski, D. A., & Brockmole, J. R. (2006). Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin & Review, 13*, 581–587. [PubMed]
- Ganel, T., Gonzalez, C. L., Valyear, K. F., Culham, J. C., Goodale, M. A., & Kohler, S. (2006). The relationship between fMRI adaptation and repetition priming. *Neuroimage, 32*, 1432–1440. [PubMed]
- Ganel, T., & Goodale, M. A. (2003). Visual control of action but not perception requires analytical processing of object shape. *Nature, 426*, 664–667. [PubMed]
- Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum.

- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635. [PubMed]
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 113–136.
- Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, *36*, 543–564. [PubMed]
- Huang, L. (2010). Visual working memory is better characterized as a distributed resource rather than discrete slots. *Journal of Vision*, *10*(14):8, 1–8, <http://www.journalofvision.org/content/10/14/8>, doi:10.1167/10.14.8. [PubMed] [Article]
- Hubel, D., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*, 215–243. [PubMed]
- Johnson, J. S., Spencer, J. P., Luck, S. J., & Schoner, G. (2008). A dynamic neural field model of visual working memory and change detection. *Psychological Science*, *20*, 568–577. [PubMed]
- Johnson, J. S., Spencer, J. P., & Schöner, G. (2009). A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. *Brain Research*, *1299*, 17–32. [PubMed]
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219. [PubMed]
- Laming, D., & Laming, J. (1992). F. Hegelmaier: On memory for the length of a line. *Psychological Research*, *54*, 233–239. [PubMed]
- Lee, B., & Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Research*, *36*, 2159–2166. [PubMed]
- Lee, D., & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, *63*, 253–257. [PubMed]
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*, 740–749. [PubMed]
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. [PubMed]
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget (2006). A Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432. [PubMed]
- Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research*, *62*, 81–92. [PubMed]
- Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, underlie object persistence. *Psychonomic Bulletin & Review*, *14*, 1199–1204.
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, *64*, 1055–1067. [PubMed]
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.
- Rolls, E. T., & Deco, G. (2010). *The noisy brain: Stochastic dynamics as a principle of brain function*. Oxford, UK: Oxford University Press.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46. [PubMed]
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*, 207–214. [PubMed]
- Tegner, J., Compte, A., & Wang, X. J. (2002). The dynamical stability of reverberatory neural circuits. *Biological Cybernetics*, *87*, 471–481. [PubMed]
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, *27*, 92–114. [PubMed]
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception & Performance*, *32*, 1436–1451. [PubMed]
- Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neuroscience*, *24*, 455–463. [PubMed]
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*, 48–64. [PubMed]
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12):11, 1120–1135, <http://www.journalofvision.org/content/4/12/11>, doi:10.1167/4.12.11. [PubMed] [Article]
- Wolters, G., & Raffone, A. (2008). Coherence and recurrency: Maintenance, control and integration in working memory. *Cognitive Processing*, *9*, 1–17. [PubMed]

- Xu, Y. (2002). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics*, *64*, 1260–1280. [[PubMed](#)]
- Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235. [[PubMed](#)]

- Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428. [[PubMed](#)]