Journal of Experimental Psychology: Learning, Memory, and Cognition

No Evidence for a Fixed Object Limit in Working Memory: Spatial Ensemble Representations Inflate Estimates of Working Memory Capacity for Complex Objects

Timothy F. Brady and George A. Alvarez

Online First Publication, November 24, 2014. http://dx.doi.org/10.1037/xlm0000075

CITATION

Brady, T. F., & Alvarez, G. A. (2014, November 24). No Evidence for a Fixed Object Limit in Working Memory: Spatial Ensemble Representations Inflate Estimates of Working Memory Capacity for Complex Objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://dx.doi.org/10.1037/xlm0000075

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/14/$12.00 http://dx.doi.org/10.1037/xlm0000075

RESEARCH REPORT

No Evidence for a Fixed Object Limit in Working Memory: Spatial Ensemble Representations Inflate Estimates of Working Memory Capacity for Complex Objects

Timothy F. Brady and George A. Alvarez
Harvard University

A central question for models of visual working memory is whether the number of objects people can remember depends on object complexity. Some influential "slot" models of working memory capacity suggest that people always represent 3–4 objects and that only the fidelity with which these objects are represented is affected by object complexity. The primary evidence supporting this claim is the finding that people can detect large changes to complex objects (consistent with remembering at least 4 individual objects), but that small changes cannot be detected (consistent with low-resolution representations). Here we show that change detection with large changes greatly overestimates individual item capacity when people can use global representations of the display to detect such changes. When the ability to use such global ensemble or texture representations is reduced, people remember individual information about only 1–2 complex objects. This finding challenges models that propose people always remember a fixed number of objects, regardless of complexity, and supports a more flexible model with an important role for spatial ensemble representations.

Keywords: visual short-term memory, ensemble perception, summary statistics, visual texture, memory capacity

Working memory is the ability to hold information actively in mind in a readily accessible state (Baddeley, 2000). The capacity of this system is severely limited, and these limits appear to have important consequences: Individual differences in working memory capacity predict differences in fluid intelligence, reading comprehension, and academic achievement (Alloway & Alloway, 2010; Daneman & Carpenter, 1980; Fukuda, Vogel, Mayr, & Awh, 2010; Oberauer, Schulze, Wilhelm, & Süβ, 2005). Thus, ever since Miller's (1956) influential proposal that working memory can store 7 ± 2 chunks of information, the study of working memory has focused on understanding the nature and units of its capacity (e.g., Brady, Konkle, & Alvarez, 2011; Cowan, 2001; Miyake & Shah, 1999). A central question has been: Is working memory really limited to storing a fixed number of chunks, independent of the content and complexity of each chunk? Or does the content of a chunk affect how many can be remembered (Simon, 1974)? This issue has been examined particularly thoroughly in the domain of visual working memory, by testing people's ability to remember objects that vary in complexity. It has been shown that when items are more complex,[1] fewer of them can be remembered with fine detail—providing strong evidence for an information limit rather than a fixed object limit in working memory (Alvarez & Cavanagh, 2004). This is important because it suggests that the fundamental limit on working memory capacity does not come from being able to store only a fixed number of objects in memory (e.g., Cowan, 2001), ruling out many classes of working memory models (e.g., models where the units are objects and these objects are stored in a fixed number of oscillatory phases; Luck & Vogel, 1998, 2013).

However, it is possible that working memory has a fixed object-limit or chunk-limit that constrains the number of items remembered (as in Miller, 1956) *and* a fixed information-limit that constrains memory fidelity (as in Alvarez & Cavanagh, 2004). In particular, although only one to two complex objects can be stored with fine detail, people might always remember *something* about a fixed number of approximately four objects, regardless of object complexity. This would suggest two limits in memory: both a limit on some shared resource and a limit on the number of objects about which any information at all can be stored.

To test this hybrid-model account, Awh et al. (2007) asked participants to remember objects and then to detect either within-

Timothy F. Brady and George A. Alvarez, Department of Psychology, Harvard University.

This work was supported by National Science Foundation CAREER Grant BCS-0953730 to George A. Alvarez.

Correspondence concerning this article should be addressed to Timothy F. Brady, Department of Psychology, Harvard University, 33 Kirkland Street, Room 702, Cambridge, MA 02138. E-mail: tbrady@wjh.harvard.edu

[1] As defined by the difficulty of perceptually processing them in a visual search task (Alvarez & Cavanagh, 2004).

category changes (e.g., a cube to another cube), which required high-fidelity representations, or cross-category changes (e.g., a cube to a Chinese character), which required only low-fidelity representations. They found that performance on within-category changes was low and consistent with remembering only one complex object (per Alvarez & Cavanagh, 2004), but that performance was much greater with cross-category changes, consistent with storage of up to four complex items. Awh et al. (2007) suggested that the higher capacity estimate for cross-category changes indicated that the true storage capacity is about four objects, even for complex objects. If this account is correct, the data of Awh et al. (2007) provide an important piece of evidence in favor of chunk or slot models of memory, suggesting a way in which fixed-chunk models can be preserved even when information limits also exist (see also Barton, Ester, & Awh, 2009; Fukuda, Vogel, et al., 2010; Scolari, Vogel, & Awh, 2008). However, one key assumption of these studies is that change detection is supported solely by the maintenance of individual items in working memory, with no contribution from other kinds of memory representations (e.g., the storage of ensembles; Alvarez, 2011; Brady & Alvarez, 2011).

Here we show results that undermine the primary evidence in favor of such hybrid models that have both a fixed-object limit and an information limit. Specifically, we show that cross-category changes severely inflate estimates of individual item capacity, particularly in displays where items form large, distinct clusters (e.g., those used by Awh et al., 2007). In such displays, cross-category changes are easier to detect because they disrupt the texture of items that cluster in the display. For example, cubes are darker than Chinese characters, and so if all of the cubes on a display are clustered, that area of the display will appear visually darker and this texture property can be remembered independently of any information about the particular items located there. Limiting the use of these texture/spatial ensemble representations reveals a true item capacity of only one to two complex objects, as opposed to the four-object limit found by Awh et al. (2007).

Our findings thus invalidate slot and chunk models where the number of objects remembered is fixed regardless of object complexity, and support a more fluid, information-limited model of the number of items stored in working memory. In particular, our results suggest only one to two complex objects can be stored. However, they do not speak to the possible presence of an upper bound on the number of objects that can be stored in memory, which is an assumption of other slot-based models (for a discussion, see Suchow, Fougnie, Brady, & Alvarez, 2014). Our results also show the importance of taking into account ensemble representations when characterizing working memory, rather than treating all information participants remember as though it arises solely from individual object representations.

Experiment 1: Replication of Awh et al. (2007)

In Experiment 1, we replicated Awh et al.'s (2007) finding that people are better at detecting cross-category changes than within-category changes when remembering complex objects. However, we conducted an additional analysis that suggests performance on cross-category changes relies on ensemble or texture representations, rather than individual item representations. To enable us to conduct this analysis, we collected data from a large number of participants, and all participants were shown exactly the same visual displays, rather than randomly generated displays for each observer. This minor change to the paradigm enabled us to examine which displays result in the best performance, and to determine the role that texture and ensemble representations play in change detection.

Method

One hundred participants performed the working memory task of Awh et al. (2007), with displays of eight complex objects. After giving informed consent, participants were tested on the Internet via Amazon Mechanical Turk and were paid $0.50 for approximately 5 min of their time. Turk users form a representative subset of adults in the United States (Berinsky, Huber, & Lenz, 2012; Buhrmester, Kwang, & Gosling, 2011), and data from Turk is known to closely match data from the lab on working memory tasks (Brady & Alvarez, 2011; Brady & Tenenbaum, 2013).

The displays consisted of intermixed cubes and Chinese characters, and participants were asked to detect changes both within-category (cube to another cube, or character to another character) and cross-category (cube to character, or vice versa; see Figure 1) on different interleaved trials. We showed all participants exactly the same individual displays, enabling us to visualize displays sorted by performance (see Figure 2), and to analyze which arrangements of cubes and characters caused the best performance in the task.

Forty-eight displays were generated in the same manner as Awh et al. (2007). In particular, we used the six 3–D cubes and six Chinese characters from Alvarez and Cavanagh (2004), and randomly selected eight stimuli from this set of 12 possible stimuli without replacement for each trial. We then randomly positioned these stimuli with the constraint that two items landed in each quadrant. Displays were presented to participants for 1,000 ms, and then after a 1,000-ms blank, a single item reappeared and participants had to report whether that item was the same or different from what it had been in the initial display. The test item appeared in the same spatial location as the item being tested, making location the cue to participants about which object was being tested. No objects were repeated in the study display, but the probe item could be the same as a different item from the study display. Thus, on some trials participants needed to use location-bound information to decide whether the probe was the same or different. This use of location as a cue is standard in the literature on visual working memory (Luck & Vogel, 1997; Zhang & Luck, 2008) and was identical to the need for location in the original study of Awh et al. (2007).

The 48 displays were presented twice to each participant—once as a *same* trial and once as a *different* trial. Thus, participants performed 96 trials total, plus two practice trials (with set size two). For half of the displays, participants were tested with a cross-category change, and for the other half of the displays, they were tested with a within-category change. Because these trial types were randomly intermixed, participants did not know in advance which kind of change detection would be required and so could not strategically choose to encode the display differently for these kinds of trials. The displays were presented in a different random order for each participant.

Different items from the display were tested for each different participant. Thus, while the initial display was the same for each

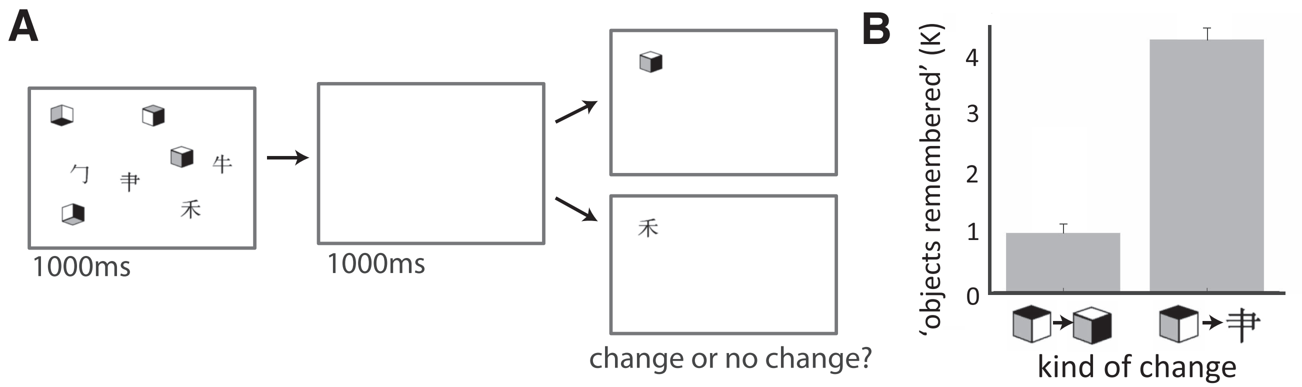

Figure 1. A: Participants performed a change detection task with either within-category (small) or cross-category (large) changes. All participants saw exactly the same initial displays but were tested on different items. B: Replicating Awh et al. (2007), participants seemed to remember only one object well enough to detect within-category changes, but four objects with sufficient fidelity to detect large changes. Error bars correspond to ±1 standard error of the mean.

participant, the tested item varied. This randomization controls for the possibility that some items might be preferentially encoded, and allowed us to assess performance for the display as a whole.

Data analysis. To enable a direct comparison with Awh et al. (2007), we estimated the number of individual items remembered using Cowan's K (Cowan, 2001): $K = (H - FA) * N$, where K is the number of items stored, H is the hit rate, FA is the false alarm rate, and N is the number of items presented.

Results

An analysis of the conditions collapsing over all displays replicated Awh et al. (2007). We found a capacity of approximately one object in the within-category change condition and approximately four objects in the cross-category change condition ($K = 0.95$ vs. $K = 4.23$)—a large and significant difference, $t(99) = 19.9$, $p < .001$.

However, we also found substantial differences in performance for different displays. Visualizing these differences reveals that participants performed best on the displays where the cubes and/or characters clustered together spatially (see Figure 2). To provide a formalization of the degree of clustering, we computed the "dispersion" of items within a category—which we formalized based on data from a pilot experiment as how far apart on the display the nearest two cubes or characters were to each other (in pixels). This dispersion measure was a major predictor of success on cross-category change detection but not within-category change detection (see regression below). This particular dispersion measure should not be taken as a cognitive model of how displays are represented—participants' representations are likely to be spatially rich, more closely resembling a visual texture representation of the display, as in Brady and Tenenbaum (2013) or Alvarez and Oliva (2009)—but this measure does provide a simple formalization of clustering and captures reliable differences in performance across displays.

We quantified the effect of dispersion on individual change-detection trials using a logistic regression model with two z-scored predictors (1) the dispersion of the display, defined as the spatial distance in pixels between the nearest two cubes or nearest two characters, whichever was closer together, and (2) whether the changed object started off as a cube or character. We found that the within-category changes were affected only by the category of the object (cubes were harder than characters, $\beta = 0.08$, $p = .06$), but there

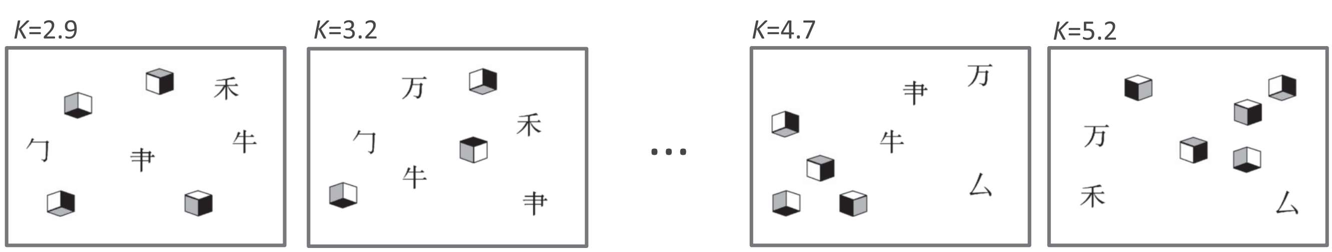

Figure 2. The individual displays where participants performed best on the cross-category changes (right) were those in which the cubes were clustered together, such that participants could detect a change from a cube to a character based on a change in clustering (an ensemble or texture representation) rather than an individual item memory. The figure shows the two worst displays (left) and two best displays (right) for illustration of this effect, with the capacity estimate for cross-category changes (K) for each display listed above it. Note that even the worst displays still contain fairly significant ensemble information, since they have only two kinds of objects present.

was no effect of dispersion ($\beta = 0.00$, $p > .10$), whereas for cross-category changes only the dispersion between objects of the same category was relevant ($\beta = -0.30$, $p < .001$; greater dispersion leads to lower capacity), with no effect of category of initial item ($\beta = -0.01$, $p > .10$). In addition to predicting individual trials, we could also look at how well the dispersion measure predicted participants' average capacity for individual displays. We found that the dispersion measure was significantly correlated with the capacity derived for each display from cross-category changes ($r = -.67$, $p < .0001$) but not within-category changes ($r = .13$, $p = .54$). Using a test that took into account the dependence between the two correlations (Steiger, 1980), the difference in these correlation values was significant ($z = -3.2$, $p < .001$), consistent with our prediction that the dispersion affects the difficulty of cross-category changes but not within-category changes.

The dispersion effect was quite large. In fact, when performance on cross-category changes was extrapolated to displays where the items were almost totally dispersed (*z* score of 3.0), the trial-by-trial regression predicted that participants should remember only 2.7 objects, even for the cross-category changes. If the only factor limiting performance were the size of the change, then performance would be identical across all displays for a given change size. Thus, the strong influence of the dispersion of the objects on performance (and the non-model-based visualization showing the effects of displays on performance, as in Figure 2) suggests that a representation of multiple items as an ensemble (e.g., a texture or spatial ensemble representation) is responsible for the improved performance on cross-category changes.

The effect of dispersion could not have been due to simple perceptual grouping (e.g., treating all of the cubes as a single unit in perception and memory, as in Anderson, Vogel, & Awh, 2013) because participants continued to represent individual details of cubes and characters, as indicated by performance in the within-category condition. If all participants had stored in memory was a representation of where the cubes on the display were (e.g., whether, for each position, the object there was a "dark" object like a cube or a "light" object like a Chinese character), this would not have allowed them to successfully answer questions about which individual cubes or Chinese characters were present within this group. Instead, participants appear to have had access simultaneously to some individual items and to ensemble properties of the display (e.g., Brady & Alvarez, 2011; Brady & Tenenbaum, 2013). However, the kind of ensemble representation present here was quite different from that in previous studies showing effects of statistical summary representations on working memory for individual items (e.g., Brady & Alvarez, 2011). In this case, rather than participants having a single ensemble representation like the mean size of the items on the display, they appear to have been representing a spatial ensemble or texture representation (e.g., where the darkest parts of the display were, or the spatial frequency distribution across the display, or other texture properties). In line with Alvarez and Oliva (2009) and Brady and Tenenbaum (2013), we call these spatial ensemble representations, since they are spatially rich texture representations, not simply point estimates of the display mean or variance (e.g., Ariely, 2001; Chong & Treisman, 2003; Haberman & Whitney, 2007).

Experiment 2: Heterogeneous Displays

To verify that spatial ensemble representations inflate capacity estimates for individual items, we ran a follow-up experiment with more heterogeneous displays that reduced the usefulness of such spatial ensemble representations. Specifically, participants were required to detect both within- and cross-category changes in displays composed of not only cubes and characters, but also of gray polygons and Snodgrass objects (other categories used by Alvarez & Cavanagh, 2004). Increasing the heterogeneity of the objects reduces the ability to form spatial ensemble or texture representations across items (e.g., store a separate representation of where the darkest areas of the display are). Thus, if this form of spatial ensemble representation inflates estimates of capacity for complex objects, then capacity estimates should be lower in these heterogeneous displays (on the order of 2–2.5 objects, according to the regression analysis of Experiment 1). However, if participants always remember four objects regardless of complexity, as claimed by Awh et al. (2007), then capacity should be four objects for cross-category changes even in these heterogeneous displays (because we used the same large changes as in Experiment 1).

Method

One hundred new participants participated in Experiment 2A, and another 100 in Experiment 2B. All methods were identical to those in Experiment 1, except that in Experiment 2A, the displays consisted of two cubes, two Chinese characters, two Snodgrass objects, and two gray polygons (set size eight). In Experiment 2B, the displays consisted of only one object of each kind (set size four). In Experiment 2B, we made the additional modification that, for cross-category changes, cubes always changed to Chinese characters (and vice versa), whereas Snodgrass objects always changed to polygons (and vice versa). This allowed us to compare Experiment 2B and Experiment 1 with exactly the same changes used, and only the surrounding context changed.

Results

The within-category condition replicated the results of Experiment 1: Taking into account only within-category changes to cubes and characters, participants remembered approximately one object (Experiment 2A: $K = 0.92$; Experiment 2B: $K = 1.21$). However, even for large, cross-category changes that required only low-fidelity representations (e.g., cube to Chinese character), participants remembered only two objects (see Figure 3). Indeed, performance here was significantly lower than in the first experiment (Experiment 2A: $K = 2.05$, compared to 4.23 in Experiment 1, $t(198) = 9.4$, $p < .001$; Experiment 2B: $K = 2.41$, $t(198) = 10.7$, $p < .001$) and in line with the performance level predicted by our dispersion model. This pattern held when examining only changes from cubes to characters or vice versa, matching the changes used in Experiment 1 (Experiment 1: $K = 4.23$; Experiment 2B: $K = 2.48$, $t(198) = 10.2$, $p < .001$). If only change size determined performance on this task, as claimed by Awh et al. (2007), then performance would have been the same across these experiments. Thus, the simple modification of making the displays more heterogeneous was sufficient to disrupt the representation participants in Experiment 1 used to detect cross-category changes, but not

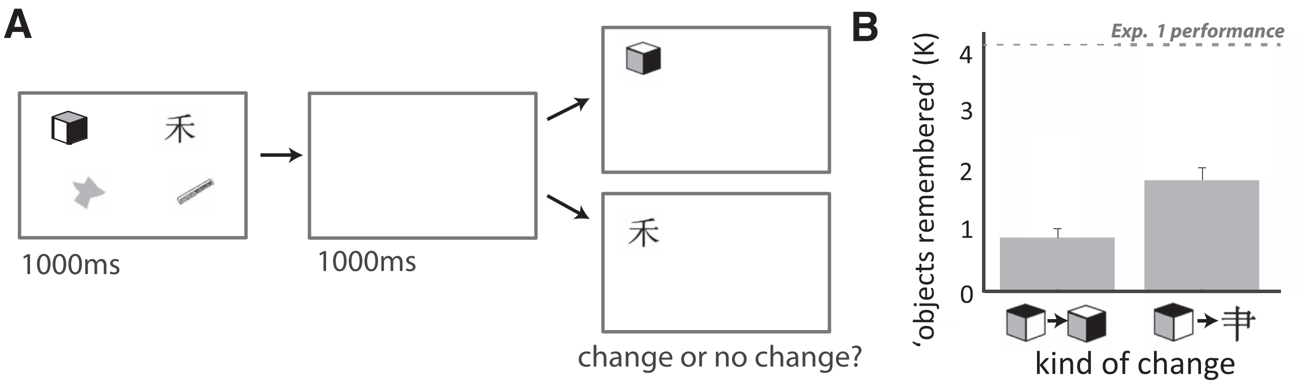

Figure 3. (A) In Experiment 2, we increased the heterogeneity of the objects in each display, minimizing clustering of items, and thus minimizing the utility of ensemble or texture representations for detecting cross-category changes. We tested both set size 4 (method shown) and set size 8. (B) Capacity was much lower for cross-category changes in Experiment 2 relative to Experiment 1, at both set size 4 and 8 (data are shown collapsed across set size). Thus, the high capacity estimate for cross-category changes in Experiment 1 (dashed line) appears to be inflated by the availability and use of texture/ensemble representations, which are unavailable in Experiment 2.

within-category changes. These results demonstrate that memory does not always hold four individual items regardless of complexity. Instead, they suggest that the capacity for complex objects is less than four, but that participants can use a global representation of spatial ensemble or texture properties of the display to detect large changes that disrupt the layout of items. When this texture representation is made less useful by increasing the heterogeneity of the displays, the ability to detect large changes is substantially reduced.

Individual differences in our experiments also support this conclusion. In particular, Awh et al. (2007) found little correlation between the ability to detect within-category changes and cross-category changes. Based on this lack of correlation, they suggested that the number of items stored (revealed by large changes) is independent of the resources used for memory fidelity (revealed by small changes). However, the current framework suggests an alternative interpretation of such results: Large changes rely on both individual item representations and spatial ensemble representations, whereas small changes primarily rely on individual item representations. As such, the theory of Awh et al. (2007) makes a different prediction from the ensemble theory for Experiment 2: In particular, only the ensemble theory predicts that performance with heterogeneous displays will be supported primarily by individual object representations for both large and small changes.

Individual differences in our data are consistent with the ensemble theory predictions. In Experiment 1, where the detection of large changes and small changes should have relied primarily on different representations (spatial ensembles and individual items, respectively), we found a relatively small correlation of $r = .26$ ($r^2 = .068$, $p = .01$) between performances on within-category and cross-category changes, in line with the finding of Awh et al. (2007). By contrast, in Experiment 2A, which has the same set size and differs only in having heterogeneous displays (which should cause detection of both large and small changes to more consistently rely on individual object representations), this correlation is much larger ($r = .55$; $r^2 = .30$, $p < .001$). Thus, detecting small and large changes is more strongly correlated within heterogeneous displays (Experiment 2A) than within homogeneous displays (Experiment 1; $z = 2.45$, $p = .01$).[2] Our finding that within-category change detection predicts nearly five times the variance in cross-category change detection in Experiment 2A compared to Experiment 1 provides further support for the idea that ensemble representations are a more important part of performance in cross-category changes in Experiment 1 than Experiment 2.

Importantly, we found that performance was nearly identical between Experiments 1 and 2A for within-category changes ($K = 0.95$ vs. $K = 0.92$; these were the two experiments with eight items present). Thus, it is unlikely that participants used a grouping strategy in Experiment 1, and an item-based strategy in Experiment 2, because such a strategy would likely result in a trade-off between memory for individual items and memory for cross-item information. Under such an account, we would expect participants to remember more individual items in Experiment 2A than they did in Experiment 1, which is inconsistent with our data. Instead, we found that the same amount of individual item information was remembered in both experiments, with only the cross-category changes affected by our heterogeneity manipulation. This is consistent with the idea that the cross-category change detection relied on an ensemble or texture memory that was disrupted by display heterogeneity.

[2] Reliabilities in the two experiments were very similar, with both experiments having relatively limited reliability due to the small number of trials performed by each participant. In particular, the reliabilities of the experiments place a limit on the maximum observable correlation between within- and cross-category capacity of approximately $r = .50$ (per Nunnally, 1970); thus, the correlation observed in Experiment 2A is about as large as could be expected, given the reliability of our estimates. The correlation in Experiment 2B, while not directly comparable to the others due to the different set size, was $r = .40$ ($p < .0001$).

Experiment 3: Comparison Between Complex and Simple Objects

Participants in Experiment 2 remembered considerably less than the fixed four object capacity argued by Awh et al. (2007) and others to represent the capacity of visual working memory and considerably less than they remembered in the homogeneous displays of Experiment 1. This provides strong evidence that spatial ensemble representations lead to an overestimate of how many individual complex objects can be remembered. However, it does not provide a direct test of another point raised by Awh et al., which is whether participants remember fewer complex objects than simple objects. Thus, in Experiment 3, we replicated Experiment 2A but added a color change detection condition to examine the capacity of the same participants to remember simple stimuli.

Method

One hundred new participants completed Experiment 3, which was a replication of Experiment 2A with an additional color change detection condition added. This condition was based on Luck and Vogel (1997) and allowed us to compare capacity for complex objects with capacity for colors. The methods and timing for the color change detection task were the same as those for the change detection task with complex objects, except that the stimuli consisted of the eight colored squares rather than eight complex objects. Following Luck and Vogel, the colors were chosen from a set of seven categorically unique colors with replacement. All trials were randomly generated and distinct for each participant. Participants performed 96 trials of the color change detection task in addition to the 96 trials of complex object change detection. The order of the color change detection task and the complex objects change detection task were counterbalanced across participants.

Results

We successfully replicated Experiment 2A's results with complex objects, with participants remembering approximately one object in the within-category change detection and 2–2.5 objects in the cross-category change detection conditions ($K = 0.71$, $K = 2.54$, respectively). The capacity of 2.5 objects in the cross-category condition was again significantly less than the capacity for complex objects observed in the cross-category condition in the cube- and character-only displays of Experiment 1, $t(198) = 7.72$, $p < .0001$. In addition, we found a capacity of $K = 4.0$ colors, significantly larger than the capacity estimate with cross-category changes in the present experiment, $t(99) = 12.58$, $p < .0001$, but not significantly different from capacity on cross-category changes in Experiment 1 ($K = 4.2$), $t(198) = 1.07$, $p = .28$. Thus, our results replicated the standard finding of a capacity of four colors and, in line with Awh et al. (2007), replicated the finding that this capacity is approximately the same as the capacity for cross-category changes in the cube- and character-only displays of Experiment 1. However, we again found a significantly reduced capacity for cross-category changes in heterogeneous displays compared to the homogeneous displays of Experiment 1 and Awh et al., suggesting that spatial ensemble/texture representations play a significant role in inflating capacity estimates for complex objects in the displays consisting of only cubes and characters.

In line with the broader literature, we used the standard color change detection task of Luck and Vogel (1997) and quantified performance in terms of the number of individual items remembered (i.e., about four objects). However, it is unlikely that performance on any working memory task is driven solely by individual item representations: Ensemble processing contributes to memory, even for displays of colored dots, where the test stimulus is a simple colored square changing to another colored square (e.g., Brady & Tenenbaum, 2013). Of course, our heterogeneous complex object displays and color displays did not have the large spatial inhomogeneity that made spatial ensemble representations so exceptionally useful in Experiment 1. Nevertheless, it is unlikely that these displays completely eliminated the impact of spatial ensembles on performance. Thus, it is likely that at least a portion of the four colors remembered and a portion of the 2.5 complex objects remembered were actually accounted for by ensemble representations.

Individual differences in Experiment 3 show a relationship between color change detection and cross-category change detection (replicating Awh et al., 2007). In particular, we found that both color change detection and within-category change detection independently explain a portion of the variance in the cross-category changes, with color change detection explaining more of the variance in a multiple regression but both playing a significant role (color $\beta = 0.88$, $p < .0001$; within-category $\beta = 0.26$, $p = .05$). This is consistent with an account where both color changes and cross-category changes rely on some ensemble processing and some individual item processing, whereas within-category changes provide a relatively direct measure of individual item processing.

In summary, when quantifying item capacity, as is generally done in the literature (e.g., Cowan, 2001; Luck & Vogel, 1997), we found that fewer complex objects can be stored than simple objects. However, the present results suggest that it important to consider, and ultimately account for, the contribution of ensemble representations to memory for both simple and complex objects. In the present experiments, ensemble representations clearly inflated capacity estimates for complex objects, and we cannot rule out the possibility that they also inflated estimates of item capacity in standard working memory tasks using within-category changes or simple color changes.

General Discussion

We found that people can remember individual-item information about only approximately one to two complex objects. This finding directly contradicts variants of the slot model which predict that people always represent information about four objects regardless of complexity, or that only fidelity is affected by object complexity (e.g., Awh et al. 2007; Barton, Ester, & Awh, 2009; Fukuda, Vogel, et al., 2010; Scolari, Vogel, & Awh, 2008). Thus, contrary to the hybrid model proposed by Awh et al. (2007), the only way a slot or chunk model can account for the full breadth of working memory data is to allow for complex objects to be split into multiple slots or chunks (Vogel, Woodman, & Luck, 2001; Xu, 2002). This limited capacity for complex objects adds to the growing list of complications for slot model accounts of visual working memory. For example, slot models must also allow for multiple copies of a single feature to be stored to increase precision (Zhang & Luck, 2008), and must allow for considerable variability in the number or fidelity of slots from trial to trial, even within an observer (e.g., Fougnie, Suchow, & Alvarez, 2012). While such modifications enable slot models to account for the broader set of findings in the literature, they begin to strain the slot concept. More generally, the present findings call attention to the prevalence of ensemble and texture representations in working memory, and underscore the need to take these representations into account in understanding the structure and capacity of working memory (see Brady, Konkle, & Alvarez, 2011, for a review).

The current results are ambiguous about whether fidelity limitations or comparison errors contribute in any way to limiting performance on within-category change detection with these stimuli. In Experiment 2, we increased the heterogeneity of the displays and found that the benefit for cross-category changes relative to within-category changes was considerably diminished. However, participants still showed some benefit of the larger changes. One possibility is that this reflects the greater difficulty of the change detection task in within-category changes (e.g., cross-category changes may be ameliorating comparison errors, since they require lower fidelity). However, we also cannot rule out the idea that some form of spatial ensemble representation was still present and allowed participants to succeed on the cross-category changes but not the within-category changes. Increasing the heterogeneity of the display is unlikely to *completely* eliminate the benefit that participants seem to accrue from storing a separate ensemble or texture representation.

Our experiments allowed for the possibility that the probe item could match other items from the study display even when it did not match the item at the same location (following Awh et al., 2007). This means that on a proportion of trials, participants needed to know not only what objects were present, but also where they were located. This requirement for location storage is common in the visual working memory literature (e.g., the work of Awh et al., 2007; Luck & Vogel, 1997; and Zhang & Luck, 2008, all have tasks that require location bindings to be stored, since items can be repeated in the same display), and the majority of the articles that claim a fixed number of items are stored actually assess not only the number of objects stored, but the number of objects stored with location-binding information. This is important because it is in theory possible that participants remember a fixed number of objects even in the cross-category change conditions, but that both Awh et al. (2007) and our experiments underestimate participants' capacity because of the need for location binding on a subset of the trials.[3]

Our experiments present an opportunity to examine whether this need for location binding provides a major constraint on performance for cross-category changes. To assess this possibility, we compared performance on cross-category change trials where probes matched other items from the display with trials where the probes were entirely unique. We found no reliable differences in performance in the experiments where there were sufficient numbers of repeated-probe trials to examine (Experiments 1, 2A, and 3). The benefit from the probe being unique in these experiments was, in terms of capacity (K), -0.01, 0.27, and -0.04, respectively; none of these differences were significant, with all ps $> .10$. Thus, participants performed the task the same way and achieved the same levels of performance on repeated-probe and unique-probe trials, which provides evidence against the idea that location binding was a major constraint in this task. Even on trials with unique probes, where location information is not required, participants still remembered only approximately two complex objects.

The current results are relevant to understanding the relationship between working memory and cognitive function more broadly. For example, it has been shown that individual differences in detecting small changes are relatively uncorrelated with individual differences in detecting large changes (Awh et al., 2007), and that the ability to detect large changes, as in cross-category change conditions, is strongly related to fluid intelligence (Fukuda, Vogel, et al., 2010). Our data suggest that cross-category change detection taps into spatial ensemble/texture processing, rather than individual item capacity. This is true for complex objects as well as for detecting large changes in color (Brady & Tenenbaum, 2013, Experiment 2), which also seem to rely on texture/ensemble representations. This raises the possibility that the correlation between working memory and intelligence found by Fukuda, Vogel, et al. (2010) only for large changes actually reflects people's ability to strategically take advantage of ensemble statistics and texture representations, rather than the ability to remember information about many individual items at once. This is an important possibility that should be explored in future research.

Moving forward, it is important to consider whether it is possible to dissociate item-level and ensemble-level representations in working memory, given that previous work has shown that these levels are integrated and that both levels of representation influence behavior (Brady & Alvarez, 2011; Orhan & Jacobs, 2013), both in cases where ensemble representations are simple summary statistics of the display, like mean size (Brady & Alvarez, 2011), and in cases like the current data, where ensemble representations seem to be more like a global texture pattern (as in Brady & Tenenbaum, 2013; see also Alvarez & Oliva, 2009; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Victor & Conte, 2004). One possibility is that functional imaging could provide the tools to separately measure individual item representations and ensemble/texture representations. This idea is compatible with evidence from functional magnetic resonance imaging (Xu & Chun, 2006) and electroencephalographs (Gao et al., 2009; Luria, Sessa, Gotler, Jolicoeur, & Dell'Acqua, 2010), examining the representation of complex objects (multi-part shapes and polygons, respectively). These studies indicate that neural markers of working memory in the lateral occipital complex and the contralateral delay activity both reach saturation after one to two complex objects are encoded, significantly lower than the saturation point of three to four simple objects (although this conclusion may not generalize to all kinds of complex objects in all settings; Fukuda, Awh, & Vogel, 2010). These neuroimaging signals are sensitive to perceptual grouping of multiple objects into a single perceptual unit (e.g., Anderson et al., 2013; Xu & Chun, 2007), but they may index only the number of individual perceptual units encoded, as opposed to the use of spatial ensemble/texture representations. The spatial ensemble or texture representations themselves are likely represented in other regions, such as the parahippocampal gyrus (Cant & Xu, 2012). Thus, neural measures may enable independent estimates of individual item representations and spatial ensemble representations in working memory, providing

[3] We thank Nelson Cowan for suggesting this possibility.

powerful tools for isolating different kinds of working memory representations.

In summary, working memory appears to store structured representations, including both individual item representations and spatial ensemble/texture representations. To accurately assess the role of "chunk size" or "item complexity" on working memory capacity, it is necessary to take both levels of representation into account. When doing so, we find that memory for complex objects is not fixed or limited only by fidelity, as predicted by fixed-object slot models of working memory. In order for visual working memory performance for all display types to be described in terms of a fixed limit of any kind—slots, chunks, or information—the field requires more sophisticated measurement and models that can capture and quantify the richly structured nature of working memory representations.

References

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106,* 20–29. http://dx.doi.org/10.1016/j.jecp.2009.11.003

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15,* 122–131. http://dx.doi.org/10.1016/j.tics.2011.01.003

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. http://dx.doi.org/10.1111/j.0963-7214.2004.01502006.x

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics: Efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences, USA, 106,* 7345–7350. http://dx.doi.org/10.1073/pnas.0808981106

Anderson, D. E., Vogel, E. K., & Awh, E. (2013). Selection and storage of perceptual groups is constrained by a discrete resource in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 824–835. http://dx.doi.org/10.1037/a0030094

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162. http://dx.doi.org/10.1111/1467-9280.00327

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. http://dx.doi.org/10.1111/j.1467-9280.2007.01949.x

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. http://dx.doi.org/10.1016/S1364-6613(00)01538-2

Barton, B., Ester, E., & Awh, E. (2009). Discrete resource allocation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1359–1367. http://dx.doi.org/10.1037/a0015792

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20,* 351–368. http://dx.doi.org/10.1093/pan/mpr057

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science, 22,* 384–392. http://dx.doi.org/10.1177/0956797610397956

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): No. 4. http://dx.doi.org/10.1167/11.5.4

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review, 120,* 85–109. http://dx.doi.org/10.1037/a0030779

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6,* 3–5. http://dx.doi.org/10.1177/1745691610393980

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human anterior-medial ventral visual cortex. *Journal of Neuroscience, 32,* 7685–7700. http://dx.doi.org/10.1523/JNEUROSCI.3325-11.2012

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43,* 393–404. http://dx.doi.org/10.1016/S0042-6989(02)00596-5

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. http://dx.doi.org/10.1017/S0140525X01003922

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466. http://dx.doi.org/10.1016/S0022-5371(80)90312-6

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications, 3*: No. 1229. http://dx.doi.org/10.1038/ncomms2237

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology, 20,* 177–182. http://dx.doi.org/10.1016/j.conb.2010.03.005

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review, 17,* 673–679. http://dx.doi.org/10.3758/17.5.673

Gao, Z., Li, J., Liang, J., Chen, H., Yin, J., & Shen, M. (2009). Storing fine detailed information in visual working memory—Evidence from event-related potentials. *Journal of Vision, 9*(7): No. 17. http://dx.doi.org/10.1167/9.7.17

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17,* R751–R753. http://dx.doi.org/10.1016/j.cub.2007.06.039

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. http://dx.doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (1998). Response from Luck and Vogel. *Trends in Cognitive Sciences, 2,* 78–79. http://dx.doi.org/10.1016/S1364-6613(98)01143-7

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17,* 391–400. http://dx.doi.org/10.1016/j.tics.2013.06.006

Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience, 22,* 496–512. http://dx.doi.org/10.1162/jocn.2009.21214

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97. http://dx.doi.org/10.1037/h0043158

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139174909

Nunnally, J., Jr. (1970). *Introduction to psychological measurement*. New York, NY: McGraw-Hill.

Oberauer, K., Schulze, R., Wilhelm, O., & Süβ, H. (2005). Working memory and intelligence—Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 61–65. http://dx.doi.org/10.1037/0033-2909.131.1.61

Orhan, A. E., & Jacobs, R. A. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychological Review, 120,* 297–328. http://dx.doi.org/10.1037/a0031541

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision, 12*(4): No. 14. http://dx.doi.org/10.1167/12.4.14

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review, 15,* 215–222. http://dx.doi.org/10.3758/PBR.15.1.215

Simon, H. A. (1974, February 8). How big is a chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science, 183,* 482–488. http://dx.doi.org/10.1126/science.183.4124.482

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Suchow, J., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. Advance online publication. *Attention, Perception, & Psychophysics*. http://dx.doi.org/10.3758/s13414-014-0690-7

Victor, J. D., & Conte, M. M. (2004). Visual working memory for image statistics. *Vision Research, 44,* 541–556. http://dx.doi.org/10.1016/j.visres.2003.11.001

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114.

Xu, Y. (2002). Limitations in object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 458–468. http://dx.doi.org/10.1037/0096-1523.28.2.458

Xu, Y., & Chun, M. M. (2006, March 2). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440,* 91–95. http://dx.doi.org/10.1038/nature04262

Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences, USA, 104,* 18766–18771. http://dx.doi.org/10.1073/pnas.0705618104

Zhang, W., & Luck, S. J. (2008, May 8). Discrete fixed-resolution representations in visual working memory. *Nature, 453,* 233–235. http://dx.doi.org/10.1038/nature06860

Received April 18, 2014
Revision received September 2, 2014
Accepted September 8, 2014 ■