Discrete resource allocation in visual working memory

Brian Barton, **Edward F. Ester**, and **Edward Awh**
Department of Psychology, University of Oregon.

Abstract

Are resources in visual working memory allocated in a continuous or a discrete fashion? On the one hand, flexible resource models suggest that capacity is determined by a central resource pool that can be flexibly divided such that items of greater complexity receive a larger share of resources. On the other hand, if capacity in working memory is defined in terms of discrete storage "slots" then observers may be able to determine which items are assigned to a slot but not how resources are divided between stored items. To test these predictions, we manipulated the relative complexity of the items to be stored while holding the number items constant. Although mnemonic resolution declined when set size increased (Experiment 1), resolution for a given item was unaffected by large variations in the complexity of the other items to be stored when set size was held constant (Experiments 2–4). Thus, resources in visual working memory are distributed in a discrete slot-based fashion, even when inter-item variations in complexity motivate an asymmetrical division of resources across items.

Keywords

working memory; resource allocation; slots; resolution

Working memory holds information in a rapidly accessible and easily updated state. Although this system is critical for virtually all forms of "on-line" cognitive processing, it is widely acknowledged that it is subject to a capacity limit of about 3–4 items (e.g., Alvarez & Cavanagh, 2004; Cowan, 2000; Luck & Vogel, 1997, Pashler, 1988; Sperling, 1960). This limit motivates an interest in how mnemonic resources are allocated when perfect retention is not possible. The present research contrasts two possible models of resource allocation in visual working memory, with a focus on the granularity with which resources are divided between the stored items. On the one hand, flexible resource models suggest that capacity in working memory is determined by a shared pool of resources that can be flexibly divided between items, such that objects of greater complexity or importance receive a larger proportion of a shared pool of resources (e.g., Alvarez & Cavanagh, 2004, Eng, Chen & Jiang, 2005, Wilken & Ma, 2004). On the other hand, capacity in working memory could be defined in terms of a limited number of discrete "slots", each capable of storing an individuated item (e.g., Rouder, Morey, Cowan, Zwilling, Morey & Pratte, 2008; Zhang & Luck, 2008). In this case, resources might be allocated in a discrete fashion, such that each item receives an equivalent proportion of resources (i.e., a single slot) regardless of variations in the priority or complexity of the items

to be stored. Thus, our goal was to test whether resources in visual working memory are allocated in a continuous fashion that is sensitive to the relative complexity of the items to be stored, or whether resources are distributed in a quantized fashion such that each stored item receives an equivalent proportion of resources, regardless of variations in complexity across items.

Experiment 1 used a change detection procedure to examine the relationship between the number of items to be stored in working memory and the resolution with which each item was represented. We found clear declines in mnemonic resolution as the number of items in the sample display increased. Flexible resource models provide a natural explanation for this inverse relationship between number and resolution. As the number of items increases, the total information load of the sample array increases, and each item receives a smaller proportion of resources; thus, mnemonic resolution declines. However, one difficulty with this conclusion is that the increase in information load is confounded with an increase in the number of items in the display. Given that mnemonic resolution might decline solely because more objects were represented in memory, this inverse relationship does not discriminate between flexible resource and slot-based models of resource allocation.

To distinguish between slot-based and flexible resource models of resource allocation, we manipulated information load while holding constant the number of items to be stored. In Experiments 2 and 3, we held set size at four items and manipulated total information load by changing the proportion of complex and simple items in the display. This design examined whether the relative complexity of the items to be stored influences the allocation of resources in working memory. If mnemonic resources are flexibly allocated based on the relative complexity of each object, then mnemonic resolution for the tested item should decline as the complexity of the other items increases. By contrast, if mnemonic resources are allocated in a slot-based fashion (i.e., resolution is determined solely by the number of objects in the sample display), then mnemonic resolution for the tested item should not vary with the complexity of the other items to be stored.

Experiment 1

Experiment 1 used a change detection procedure to examine whether the resolution of representations in working memory declines as the number of stored items increases. To obtain a behavioral measure of mnemonic resolution, we relied on an approach that enables separate estimates of the number and resolution of the representations in working memory (Awh, Barton & Vogel 2007). Thus, we begin with a brief description of the prior work. The Awh et al. (2007) study tested the hypothesis that the maximum number of items that can be stored in working memory declines as object complexity increases. Alvarez and Cavanagh (2004) provided support for this hypothesis by showing a near-perfect inverse relationship between object complexity and change detection performance, such that accuracy declined as complexity increased. They concluded that fewer objects can be maintained in working memory when complexity is high. Awh et al. (2007) replicated this empirical pattern, but raised the possibility that change detection accuracy declined with complex stimuli because of errors in detecting relatively small changes, rather than from a reduction in the number of items in memory. Two findings supported this hypothesis. First, there was a close correspondence between object complexity and inter-item similarity, in line with the hypothesis that comparison errors were more likely with complex items. Second, when sample-test similarity was reduced and comparison errors were minimized, change detection performance was equivalent for simple and complex objects. These findings suggest that visual working memory represents a fixed number of items, regardless of complexity.

A key point for the present research is that the change detection procedure may measure completely different aspects of memory ability, depending on the size of the changes that have to be detected. When the changes are large and comparison errors are minimized, this procedure provides a relatively pure estimate of the number of items that can be maintained in working memory (Pashler, 1988). By contrast, when the changes are relatively small, change detection performance may also be limited by the resolution of representations in memory, given that higher resolution is required to detect smaller changes. An analysis of the individual differences in the Awh et al. (2007) study supported this interpretation by revealing a two-factor model of performance. Change detection accuracy was tightly correlated across the number-limited (i.e., large change) conditions, and performance was also correlated across the resolution-limited (i.e., small change) conditions, showing that these measures were reliable indices of ability. However, performance in the number-limited conditions did not correlate with performance in the resolution-limited conditions. Thus, Awh et al. (2007) suggested that number and resolution represent distinct facets of memory ability.

Following Awh et al. (2007), the present research measured the effect of set size in a procedure that provided separate measures of performance under number-limited (i.e., large change) and resolution-limited (i.e., small change) conditions. To anticipate the results, accuracy in the resolution-limited condition declined as set size increased. This result falls in line with earlier studies that measured memory performance as a function of set size (e.g., Ericksen & Lappin, 1967; Palmer, 1990; Reicher, 1969). For example, Palmer (1990) measured the precision with which observers could remember the length of line segments over a brief delay, while manipulating set size from one to four items. He found that the length threshold difference -- the amount of change in length needed to achieve a criterion level of performance -- doubled when set size increased from one to four items. The inverse relationship between memory performance and set size is consistent with the hypothesis that mnemonic resolution declines as set size increases. However, this empirical pattern alone does not establish this point, because increased errors with larger set sizes could be caused by a higher incidence of storage failures (i.e., items that are never encoded into working memory) instead of a reduction in the resolution of representations in memory. Thus, to analyze the data from Experiment 1 we used performance in a number-limited condition to estimate the frequency of errors due to a failure to store the critical item ("missing slot" trials). This estimate of missing slots enabled a more precise estimate of how often changes were missed even when the critical item was stored. Using the incidence of such comparison errors as an operational definition of mnemonic resolution, we examined whether resolution in working memory declined as set size increased.

Method

Observers—13 subjects received course credit for one hour of participation in Experiment 1. The subjects ranged from 18–30 years of age, and had normal or corrected-to-normal vision.

Stimuli—The stimuli were shaded cubes and Chinese characters adapted from Alvarez and Cavanaugh (2004). Each object fit snugly into a square region that subtended approximately 3.3×3.3 degrees of visual angle from a viewing distance of approximately 60 cm. On each trial, two or four objects were presented in randomly selected positions, with the restriction that no more than one object could appear in each quadrant of a light gray square region subtending 30×30 degrees of visual angle.

Procedure—The sequence of events in a single trial is illustrated in Figure 1. Each trial began with the onset of a light gray square region with a central fixation point, in which stimuli could be presented. After 1,092 ms, two or four objects appeared for 500 ms, randomly selected with replacement from both categories, with the constraint that no object could appear more than twice. A 1,000 ms delay period followed, after which a single probe object was presented and

remained visible until the subject pressed the "z" key to indicate "same" or the "/" key to indicate "different." In 5 of every 11 change trials, the probe object was selected from the same category as the test object ("within-category" changes, such as a cube changing into a cube). In the remaining 6 of 11 change trials, the test item was from a different category from the sample item ("cross-category" changes, such as a Chinese character changing into a cube). Responses were unspeeded, with instructions placing high priority on accuracy. Each subject completed 8 blocks of 48 trials, with trial order randomized within each block. Each block included 24 instances of each category and set size, and changes occurred with probability .5.

Results and Discussion

Awh et al. (2007) found that performance in the cross-category condition provides an estimate of the number of items each individual can maintain, while performance in the within-category condition is limited by mnemonic resolution. Thus, separate analyses were performed for cross-category and within-category trials. In each case, change trial data were analyzed in combination with a common set of no-change trials, so that response bias (i.e., a bias towards either "change" or "no-change" responses) could be taken into account.

We note that response bias can be influenced by the difficulty of target detection. For example, in a block of trials where all the changes are very large, observers might adopt a more conservative response threshold, leading to higher accuracy in the no-change trials. Thus, in a design where large and small changes are blocked, it would be inappropriate to compare accuracy in the change trials across these conditions because differences in response thresholds could lead to differences in accuracy that are not related to mnemonic resolution per se. This raises the question of whether we have obscured such differences between the within-category and cross-category conditions by using a common set of "no-change" trials to measure performance. On the contrary, we reasoned that randomly intermixing the within-category and cross-category trials ensured that a common response threshold was in place during no-change trials. Because both cross-category and within-category changes were possible, observers had to respond to no-change trials using a common response threshold. The random intermixing of these conditions therefore ensured that differences in the detection of cross-category and within-category changes were not an artifact of differences in response bias.

Accuracy in the within-category change trials and same trials was assessed with a three-way ANOVA that included the factors set size (2 or 4), object type (shaded cubes or Chinese characters), and trial type (same or different). We found a significant main effect of object type, driven by better performance with Chinese characters (M = 0.82) than with shaded cubes (M = 0.72), $F(1,12) = 39.68$, $p < 0.001$, $\eta^2 = .07$. There was a significant main effect of trial type, $F(1,12) = 25.35$, $p < 0.001$, $\eta^2 = .24$, such that subjects performed better on same trials (M = 0.86) than different trials (M = 0.68). We also found a significant main effect of array size, $F(1,12) = 45.22$, $p < 0.001$, $\eta^2 = .16$, showing that performance was better with array size 2 (M = 0.85) than array size 4 (M = 0.69). Finally, there was a significant interaction of trial type and object type, $F(1,12) = 28.8$, $p < 0.001$, $\eta^2 = .08$, because accuracy in different trials was worse for cubes (M = .55) than for characters (M = .74) while accuracy in same trials was equal for the two types of objects (M = .86 for both cubes and characters).

To summarize the results from the within-category and no-change trials, we replicated the findings of Alvarez and Cavanagh (2004), who observed better performance with Chinese characters than with shaded cubes. In addition, we found a clear effect of set size; performance in the two-item condition was 16% higher than in the four-item condition. Given that performance in the within-category condition is resolution-limited (Awh et al., 2007), these data are consistent with the conclusion that resolution declines as set size increases. As we have noted, however, errors in this task can also occur because of a failure to store the critical item in memory. Thus, it is necessary to take into account the probability that each item was

encoded into memory to reach a firm conclusion regarding the relationship between mnemonic resolution and set size. We return to this point following our discussion of the results from the cross-category trials.

Accuracy in the cross-category change trials and no-change trials was analyzed with a three-way ANOVA, with the factors set size (2 or 4), object type (shaded cubes or Chinese characters), and trial type (same or different). There was significant main effect of array size, $F(1,12) = 19.51$, $p < 0.001$, $\eta^2 = .11$, such that accuracy was higher for set size 2 (M = 0.94) than set size 4 (M = 0.88). There was a significant main effect of trial type, $F(1,12) = 24.09$, $p < .001$, $\eta^2 = .19$, such that accuracy was higher with different trials (M = 0.96) than same trials (M = 0.86), In line with Awh et al. (2007), change detection in the cross-category trials was equivalent for Chinese characters (M = .87) and cubes (M = .89), so that there was no significant effect of object type $F(1,12) = 0.64$, $p > 0.44$. Finally, there was a significant interaction of trial type and array size, $F(1,12) = 15.95$, $p = 0.002$, $\eta^2 = .05$, reflecting the fact that accuracy in the set size 2 condition was similar for same trials (M = .92) and different trials (M = .96), while in the set size 4 condition accuracy was lower for same trials (M = .80) than for different trials (M = .95).

To summarize the results from the cross-category and no-change trials, the cross-category condition replicated the findings of Awh et al. (2007), who found that minimizing sample-test similarity with cross-category changes led to equivalent change detection performance with cubes and Chinese characters. Using the formula developed by Pashler (1988) and refined by Cowan (2000)[1] to examine performance in the cross-category condition, we obtained capacity estimates (K) of 3.0 for both the cubes and the characters in the 4-item condition (where ceiling effects were less of a concern than in the 2 item condition).

The primary purpose of Experiment 1 was to examine the relationship between set size and mnemonic resolution. As we have noted, performance in the within-category condition is limited by mnemonic resolution, because this condition requires the detection of small changes between similar items. Nevertheless, errors can also arise in the within-category condition because of a failure to represent the critical item in memory; moreover, such errors due to "missing slots" would be more likely in the four item condition than in the two item condition. To correct for the influence of missing slots, we used performance in the cross-category condition to estimate the number of items that each subject could hold in memory. This in turn provided an estimate of how often each subject made errors due to missing slots in the within-category condition. Given an estimate of errors due to missing slots, we were able to obtain a relatively pure estimate of how often comparison errors occurred – our operational definition of mnemonic resolution – even when the critical item was stored in working memory.

Defining C, a Measure of Mnemonic Resolution—Assuming that 500 ms is enough time to fully encode these objects (see Alvarez & Cavanagh, 2004), accuracy (Acc) in the within-category condition should be equal to the probability that the object actually was encoded into working memory (P_{mem}, where P_{mem} = k/set size, and k is estimated using the cross-category and no-change trials) multiplied by the probability that the sample and test are compared correctly (C), plus a correction for guessing based on the assumption that subjects would guess correctly half of the time when the object was not stored.

$$Acc = (P_{mem} {}^* C) + (1 - P_{mem})/2$$

Solving for C:

---

[1] Where k is the number of items stored, k = set size * (hit rate – false alarm rate).

$$C=\{Acc - (1 - P_{mem})/2\}/(P_{mem})$$

This formula provides an estimate of the probability of correct comparison during the within-category condition, while correcting for errors that occur due to items that were not stored. Given that subjects were making unspeeded responses to a single probe item, we reasoned that the primary source of comparison errors in the within-category condition was the limited resolution of the stored memories. Thus, we used C as our operational definition of mnemonic resolution.

The results of this analysis are illustrated in Figure 2. A two-way ANOVA was done on estimates of C from each observer, with the factors array size (two or four) and object type (cubes or characters). We found a significant main effect of object type, $F(1,12) = 27.14$, $p < 0.001$, $\eta^2 = .24$, such that correct comparison was less likely for cubes (C = 0.75) than for Chinese characters (C = 0.89), in line with the higher sample-test similarity in the cube condition (Awh et al., 2007). There was also a significant main effect of array size, $F(1,12) = 19.34$, $p < 0.001$, $\eta^2 = .19$, such that correct comparison was less likely for an array size of 4 (C = .76 than for an array size of 2 (C = .88). Thus, mnemonic resolution declined as the number of stored items increased.

Experiment 2

The results of Experiment 1 make the important point that mnemonic resolution declines as set size increases. However, as we noted above, the decline in mnemonic resolution when set size increased from two to four does not discriminate between flexible resource and slot-based models of resource allocation. The effect of set size could be explained either by increased information load or by the increased number of objects to be stored. To test whether information load *per se* influences mnemonic resolution, it is necessary to avoid confounding information load with the number of stored objects. Thus, in Experiments 2–4 we manipulated object complexity while holding constant the number of objects in the sample display. Experiments 2 and 3 use the same Chinese characters and cubes (adapted from Alvarez & Cavanagh, 2004) that were used in Experiment 1, with a constant set size of four items. If the cubes are assumed to be more complex than the characters, then the total information load of a display will increase as the proportion of cubes in the display increases. Thus, by manipulating the proportion of cubes and characters in these four-item displays, we were able to examine how mnemonic resolution for a given item was affected by the total information load in the sample display. To further generalize this result, Experiment 4 replicated Experiments 2 and 3 using a new set of stimuli and a different procedure for measuring the clarity of the stored representations. To reiterate, flexible resource models predict that mnemonic resolution for a given item should decline as the total information load in the sample display increases, because more resources should be allocated to more complex items. By contrast, slot-based models of resource allocation suggest that mnemonic resolution for a given item is determined solely by the number of items to be stored, without regard to the relative complexity of those items.

Method

Subjects—16 subjects were paid $8.00 for one hour of participation in Experiment 2, with the same constraints as in Experiment 1.

Stimuli—The stimuli are identical to those used in Experiment 1, except that set size was held constant at 4 items and sample-test comparisons were always within-category (i.e., cubes always changed into cubes and characters always changed into characters).

Procedure—The procedure is the same as Experiment 1, except that the objects in the sample display contained either 4 of the same type of object (uniform, with four cubes or four characters) or 2 of each type of object (mixed, with two cubes and two characters).

Results and Discussion

Set size was not manipulated in this experiment. Thus, even though some errors in this task were likely due to missing slots, we used raw accuracy in change detection to test whether mnemonic resolution was influenced by object complexity. The logic of this analysis therefore relied on past observations that the number of items maintained in working memory does not vary with the complexity of the items in the sample display (e.g., Awh, Barton & Vogel, 2007; Scolari, Vogel & Awh, 2008[2]). The data were analyzed with a three-way analysis of variance including object type (shaded cubes or Chinese characters), trial type (same or different), and array type (uniform or mixed). We found a significant main effect of object type, $F(1,15) = 70.65$, $p < 0.001$, $\eta^2 = .18$, such that change detection was less accurate with shaded cubes (M = 0.65) than with Chinese characters (M = 0.77). There was a significant main effect of trial type, $F(1,15) = 16.82$, $p < 0.001$, $\eta^2 = .16$ such that accuracy on same trials (M = 0.77) was higher than on different trials (M = 0.65). There was also a significant interaction between trial type and array type, $F(1,15) = 22.83$, $p < .001$, $\eta^2 = .06$, because accuracy during same trials was higher for mixed trials (M= .80) than for uniform trials (M= .74), while accuracy during different trials was higher for uniform trials (M= .69) than for mixed trials (M = .61). There was also a significant interaction between object type and trial type, such that accuracy on different trials was lower for cubes (M=.63) than for characters (M = .78), $F(1,15) = 11.98$, $p < .01$, $\eta^2 = .04$. Finally, the critical result in this experiment is that there was no significant effect of array type, $F(1,16) = 0.35$, $p = .57$, and no interaction of array type and object type, $F(1,15) = 1.4$, $p = .25$, suggesting that resolution-limited performance with a given object (cube or character) was not affected by the complexity of the other items in the display.

Recall that flexible resource models suggest that a common pool of resources is allocated between objects, with a higher proportion of resources given to more complex objects. From this perspective, change detection with a Chinese character should be better in the uniform condition (when the character is presented with three other characters) than in the mixed condition (when the character is presented with one additional character and two cubes), because total information load is smallest in the uniform condition. In fact, change detection with Chinese characters was equivalent in the uniform (M = .77) and the mixed (M = .78) conditions, $t(15) = .74$, $p = .47$ (see Figure 3). Likewise, change detection with cubes was equivalent in the uniform (M = .66) and the mixed (M = .63) conditions, $t(15) = 1.1$, $p = .28$, even though total information load was highest in the uniform condition. These data are inconsistent with the hypothesis that more resources are allocated to more complex objects. Although we replicated previous observations that change detection is more difficult with more complex objects, the resources allocated to a given object were unaffected by the complexity of the other items in the sample array.

---

[2]The Awh et al (2007) provided a direct comparison between capacity estimates for the same Chinese characters and shaded cubes as those in the current experiment. When sample-test similarity was minimized by using cross-category changes, capacity estimates were equivalent for homogenous arrays of cubes and Chinese characters, and at the same level as for simple colored squares. Scolari et al. (2008) came to the same conclusion by comparing cross-category change detection with the shaded cubes intermixed with face stimuli and change detection with simple colors; here again, capacity estimates were equivalent when large changes minimized comparison errors.

Experiment 3

In Experiment 2, the sample arrays were presented for 500 ms. This duration should have provided ample time for encoding the four sample items[3]. It is possible, however, that flexible resource allocation requires more time than stimulus encoding alone. In this case, 500 ms may not have provided enough time to allow for flexible allocation of mnemonic resources across the four items in the sample array. In Experiment 3, we attempted to address this possibility by presenting the objects sequentially (375 ms presentations with a 125 ms inter-stimulus interval for a total presentation period of 1875 ms) to provide ample time for flexibly allocating resources across the four items in the sample display. In addition, the mixed displays in Experiment 3 contained three of one object type and one of the other object type (i.e., three cubes and one character, or one cube and three characters) so that there was a larger variation in total information load across the different experimental conditions.

Method

Subjects—17 subjects were paid $8.00 for one hour of participation in Experiment 3, with the same constraints as Experiments 1 and 2.

Stimuli—The stimuli were identical to those used in Experiment 2.

Procedure—The procedure was identical to Experiment 2, except that the stimuli were presented sequentially for 375 ms each with a 125 ms ISI (1875 ms total presentation period) and the objects presented contained either 4 of the same type of object (uniform), or one of one type and three of the other (mixed). The tested item was randomly selected on each trial, with cubes and characters tested equally often.

Results and Discussion

The data were analyzed with a three-way analysis of variance including the factors object type (shaded cubes or Chinese characters), trial type (same or different), and display type (uniform, 1 cube and 3 characters, or 3 cubes and 1 character)[4]. The results are illustrated in Figure 4. Replicating Alvarez and Cavanagh (2004), we found a significant main effect of object type, $F(1,16) = 30.36$, $p < 0.001$, $\eta^2 = .06$ such that change detection was worse with shaded cubes (M = 0.65) than with Chinese characters (M = 0.78). There was an interaction between object type and trial type, $F(1,16) = 11.88$, $p < .01$, $\eta^2 = .06$, because the difference between cube and character accuracy was larger for different trials (.18) than for same trials (no difference). Importantly, Experiment 3 replicated the findings from Experiment 2 in finding no reliable difference between performance for uniform and mixed arrays, $F(2,32) = 1.81$, p = 0.18, and no interaction between object type and array type, $F(2, 32) = 1.15$, $p = .33$[5].

To summarize the results of Experiment 3, we replicated the results of Experiment 2 while providing more time for the possibility of flexible resource allocation and a stronger variation

---

[3]For example, Alvarez and Cavanagh (2004) measured change detection with displays that contained up to 15 of these shaded cubes, and found that accuracy did not differ between exposure durations of 450 and 800 ms. This suggests that 500 ms is enough time to encode four of these cubes.

[4]The sequential presentations used in this study raise the question of whether there were effects of ordinal position. Indeed, a separate analysis revealed a reliable recency effect, with higher change detection accuracy for items that appeared in the third and fourth positions (76%) than those that appeared first or second (67%) ($p < .01$). If this were due to true increases in mnemonic resolution for the third and fourth items, it would contradict the slot-based model's prediction that resources are equally divided between all stored items. However, this recency effect could also be explained by an increased probability that the third and fourth items were stored, rather than by changes in mnemonic resolution. Thus, because serial position effects were not diagnostic with respect to primary theoretical question, our main analysis collapsed across ordinal position.

[5]Because the cubes and characters were tested equally often, there was a higher probability of testing the "singleton" item than the majority items in the heterogeneous displays. If subjects were aware of this bias, they could treat singleton status as a "cue" to pay more attention to that object. However, no evidence of bias towards the singleton objects was found with either cube or character probes.

in total information load across the uniform and mixed conditions. In both Experiments 2 and 3, significant variations in total information load had no influence on mnemonic resolution for the tested item.

Recall that it was uncertain whether increased set size or object complexity explained the inverse relationship between set size and resolution observed in Experiment 1. Experiments 2 and 3 address this ambiguity by holding set size constant and demonstrating that mnemonic resolution was unaffected by large variations in the average complexity of the stored items. Thus, taken together, Experiments 1–3 suggest that set size and not total information load determines resolution in visual working memory. This suggests that resources in visual working memory are allocated in a discrete fashion, such that observers can choose which items are assigned to slots, but are not able to bias how resources are divided between the items in memory.

Experiments 4a and 4b

Our interpretation of Experiments 2 and 3 relies on the assumption that shaded cubes are more complex than Chinese characters. However, given that object complexity (as defined by search efficiency in the Alvarez & Cavanagh (2004) study) covaried with inter-item similarity (Awh et al., 2007), and inter-item similarity can influence search efficiency (Duncan & Humphreys, 1989), there is room for doubt regarding the relative complexity of these stimuli. Thus, we attempted to replicate the results of Experiments 2 and 3 using stimuli whose relative complexity was unambiguous. In experiment 4b, observers were presented with displays containing oriented teardrop stimuli and simple or complex grids (see Figure 6). The complex grids were created by randomly filling 16 of 25 cells within a 5×5 matrix. The simple grids were created by randomly filling in 2 of 4 cells within a 2×2 matrix. Although we cannot offer a comprehensive definition of visual complexity, we reasoned that virtually any convincing definition of complexity would indicate that the 5×5 grids are more complex than the 2×2 grids.

According to a slot-based model of resource allocation, mnemonic resolution for the teardrop stimuli should be unaffected by the relative complexity of the grids present in a display. The interpretation of such a null result, however, relies on the assumption that these teardrop and grid stimuli compete for resources within a common working memory system. To verify this assumption, Experiment 4a tested whether mnemonic resolution for one of these stimuli declined when an object from the other category was added to the sample array. Because demonstrating a set size effect did not require a manipulation of grid complexity, only the simple grids were used for Experiment 4a.

Method

Subjects—9 subjects were given course credit for a one hour experimental session.

Stimuli—Two types of stimuli were presented. There were "teardrop" stimuli (illustrated in Figure 6) that varied in orientation only. The teardrops subtended 3.0×1.5 degrees of visual angle, and could appear in 36 possible orientations (starting at vertical and evenly spaced 10 degrees apart). In addition, there were simple grids that were created by randomly selecting and filling 2 of the cells within a 2×2 grid. These grids subtended 3×3 degrees of visual angle from a viewing distance of approximately 60cm. Either one or two objects were presented in each trial, with each object appearing in a separate (randomly selected) quadrant of the display (with the center of each object at 4.7 degrees of eccentricity).

Procedure—Instead of using the same change detection procedure, we used a procedure described by Eng, Chen and Jiang (2005) in which observers indicated which of two possible probe objects matched one of the objects in the sample display. This procedure had the virtue

of allowing us to manipulate the precision of the required judgment on every trial (instead of only during change trials) by controlling the size of the difference between the original object and the mismatched object.

Each trial began with the onset of a fixation point in the center of a light gray screen. After 1000 ms, either one teardrop, one grid, or one of each stimulus was presented for 200 ms. This exposure duration was chosen to preclude eye movements during the encoding phase of the trial. Each object occupied a separate and randomly selected quadrant in the display. A 1,000 ms delay period followed, after which two probe objects of the same type as one of the originally presented objects appeared on either side of the original object's location (3 degrees to the right and left of the original position). The probe objects remained visible until the subject pressed the "z" or the "/" key to indicate that the left or right probe object was the same as the original.

In two item displays, each object was equally likely to be tested in the probe display. When the teardrop was the probed object, the orientation of the mismatched teardrop was 20, 40 or 80 degrees different from the original teardrop. When a simple grid was probed, one of the two filled cells was re-positioned in the mismatched grid. Each subject completed 4 blocks of 72 trials, with trial order randomized within each block. Within each block, set size 1 and 2 were equally likely (36 trials each). Within set size 2, teardrops and grids were equally likely to be tested (18 trials each). When teardrops were tested, each change size was tested equally often (6 trials each for changes of 20, 40 and 80 degrees).

Experiment 4a Results and Discussion

Accuracy with the teardrops (illustrated in Figure 5) was analyzed with a 2-way ANOVA that included the factors set size (one teardrop alone vs. one teardrop stored with one simple grid) and the size of the mismatch between the original and the mismatched probe (20, 40 or 80 degrees). There was a clear set size effect, with higher accuracy when one teardrop was stored alone (M = .90) then when one teardrop was stored with one simple grid (M = .80) (F(1, 18) = 10.0, $p < .02$, $\eta^2 = .10$). In addition, we observed the expected effect of change size, with better performance at larger change sizes (M = .74, .91 and .91 for the 20, 40 and 80 degree changes, respectively) (F(2,16) = 32.0, $p < .001$, $\eta^2 = .32$). Accuracy with the simple grids also showed a significant set size effect, with higher accuracy when one simple grid was stored alone (M = .98) than when one simple grid was stored with a teardrop (M = .91), t(8) = 4.9, $p < .001$. These set size effects demonstrate that the teardrop and grid stimuli compete for representation in a common working memory system.

Experiment 4b

Experiment 4b tested the effect of changing grid complexity on mnemonic resolution for the orientation of the teardrops. Each display in Experiment 4b contained two grids (either simple or complex) and one teardrop. If a larger proportion of mnemonic resources are allocated towards more complex objects, then the teardrop should be remembered more precisely when it is presented with the simple grids, because more resources should be left over for the teardrop. By contrast, if resources are allocated in a discrete slot-based fashion, then mnemonic resolution for the teardrop should not be influenced by the complexity of the other items in the sample array.

Method

Subjects: 14 subjects were given course credit for one hour of participation.

Stimuli: The stimuli in Experiment 4b were the same teardrops and grids used in Experiment 4a, except that complex grids were used in addition to the simple grids. The complex grids were created by randomly selecting and filling 16 of the 25 cells within a 5×5 grid. When the

tested item was a complex grid, the mismatched grid was created by randomly re-positioning 3 of the 16 filled cells within the 5×5 grid. The overall size of the complex grids was identical to that of the simple grids (i.e., cell size was smaller in the complex grids). Each display contained one teardrop and two grids of the same complexity. Finally, an exposure duration of 500 ms was used to preclude encoding-limited performance with these more complex sample displays.

Procedure: The procedure was identical to that used in Experiment 4a, except for the composition of the sample displays which now contained three objects instead of two. From each block of 72 trials, the teardrops were equally likely to be presented with the simple and complex grids (36 trials each). Each of the three items in the sample display was equally likely to be tested during the 72 trial block, leading to 24 teardrop trials, 24 simple grid trials and 24 complex grid trials per block. For the 24 teardrop trials, 8 trials were used to test each of the three change sizes (20, 40, and 80 degree offsets).

Results and Discussion—Recall that the flexible resource and slot-based models of resource allocation made different predictions regarding the effect of grid complexity on performance with the teardrops. The predictions for performance with the grids, however, do not differ between these models because both predict that change detection will be worse with the complex grids, because of higher sample-test similarity (Awh et al, 2007). Thus, we conducted separate analyses of accuracy with the teardrops and the grids. A paired t-test of the grid trials showed that accuracy was higher with the simple grids (M = .70) than with the complex grids (M = .57), t(13) = 4.4, $p < .001$ (see Figure 7). This complexity effect replicates the results of Alvarez and Cavanagh (2004), and supports the face validity of our manipulation. The key analysis, however, focused on whether performance with the teardrops was affected by grid complexity.

Accuracy with the teardrop stimuli (illustrated in Figure 8) was analyzed with a 2-way ANOVA that included the factors complexity (high or low) and the size of the mismatch between the original and the mismatched probe (20, 40 or 80 degrees). There was a main effect of mismatch size, $F(2,26) = 4.95$, $p < .05$, $\eta^2 = .08$, with monotonic increases in accuracy as the size of the mismatch increased. The monotonic relationship between the size of the mismatch and accuracy shows that performance was limited by observers' ability to discriminate between the relatively similar pair of probes in the test phase of the trial. The key finding, however, was that accuracy with the teardrops showed no main effect of complexity, $F(1,13) = 3.10$, $p = .10$, and no interaction of complexity and mismatch size, $F(2,26) = .78$, $p = .47$. Indeed, the non-significant effect of complexity in this experiment went in the opposite direction from the predictions of the flexible resource model, with numerically lower accuracy in the simple grid condition (M = .68) than in the complex grid condition (M = .71). Thus, contrary to the predictions of the flexible resource model, we found no evidence that mnemonic resources were biased towards items of higher complexity. If this had happened, then teardrop performance should have benefited from more resources in the simple grid condition. Instead, resolution-limited performance with these teardrops was equivalent when they were stored with simple and complex grids. Thus, despite the strong influence that the number of objects has on mnemonic resolution (Experiment 1), Experiments 2–4 suggest that the precision of a given object's representation in working memory is not influenced by the complexity of the other items to be stored.

General Discussion

We attempted to distinguish between two models of resource allocation in visual working memory. On the one hand, flexible resource models suggest that this process is sensitive to the relative complexity of the items to be stored in working memory, such that more complex items

receive a larger share of resources. On the other hand, slot-based models of resource allocation suggest a more rigid allocation strategy, in which each item's share of resources depends only on the number of objects to be stored. In Experiment 1, we used a change detection procedure to examine the relationship between mnemonic resolution and the number of items in the sample display. Using an analytic approach that corrected for failures to store the critical item, we found clear drops in mnemonic resolution as the number of items in the sample display increased. This result made the important point that mnemonic resolution declines as set size increases, suggesting that fewer mnemonic resources are available for each item as the number of items increases. However, because set size was confounded with total information load, the inverse relationship between set size and resolution does not distinguish between flexible resource and slot-based models of resource allocation.

Thus, Experiments 2–4 examined conditions where it was clear that these models make different predictions. Specifically, flexible resource models predict that the complexity of a given item determines the proportion of resources that are allocated to that item. Therefore, flexible resource models predict that when the number of items to be stored is held constant, the resources that are granted to a specific item should decline as the complexity of the other items increases. Across three different experiments and two different sets of stimuli, this prediction was disconfirmed. In each case, observers' ability to carry out resolution-limited judgments for a critical item was unaffected by changes in the average complexity of the other items that had to be stored. This null effect of complexity contrasts with the effect of set size found in Experiment 1 (and Experiment 4a), in which clear declines in mnemonic resolution were observed as set size increased. Thus, we conclude that resources in visual working memory are distributed in a discrete, slot-based fashion, such that the resources devoted to a single item are determined by the number of objects that must be retained.

This slot-based model of resource allocation is consistent with recent findings from Zhang and Luck (2008). Using a procedure that provided independent estimates of the number and resolution of the representations in working memory, they found evidence for a high-threshold model of memory in which a small subset of the available items are maintained while no information is retained about other items. Directly relevant to the present findings, they examined whether observers could use a pre-cue to bias resource allocation across the stored items. Specifically, a precue indicated which of four items was most likely to be probed at the end of the trial. The rationale was that observers should – if possible – devote the largest share of resources to the cued items, leaving less available for the other items; this should lead to higher resolution for the cued item relative to the others. Instead, the results showed that while cued items were more likely to be encoded into memory, the resolution of those representations was unaffected by the cues. Thus, while our results showed that resources are not asymmetrically distributed when there are large variations in complexity across items, the Zhang and Luck (2008) findings suggest that informative precues affected only which objects were encoded rather than the precision with which those items were stored. The present findings as well as those of Zhang and Luck (2008) contradict the predictions of the flexible resource model, which suggests that observers should be able to flexibly allocate a larger proportion of resources to items of higher priority or greater complexity.

We have suggested that resources are allocated in a slot-based fashion, such that mnemonic resolution is determined only by the number of items that are stored in working memory. This suggestion, however, leaves open questions about how slot-based resource allocation is implemented. Zhang and Luck (2008) offered a more detailed hypothesis for why mnemonic resources are "quantized" at the level of objects. Specifically, these authors proposed a "slot averaging" model, according to which multiple slots can be assigned to a single item when the number of stored items is smaller than the number of available storage slots in working memory. In this case, small improvements in mnemonic resolution are predicted for sub-span

set sizes because of the benefits of redundant representation. Thus, the slot-averaging model explains why mnemonic resolution decreases as set size increases, because larger set sizes preclude using multiple slots for a single item. Moreover, the slot averaging model provides a natural explanation for why mnemonic resolution for a given item is not affected by the complexity of other items. Specifically, if the slots themselves are the basic units of mnemonic resources, then it is clear why resources cannot be re-distributed in a more fine-grained fashion when one object has higher priority or higher complexity.

Recall that Awh et al. (2007) found that there was no correlation between the number of items that an individual could store and the resolution of those representations. This result suggests that number and resolution may represent distinct aspect of memory ability. This conclusion corroborated the findings of Xu and Chun (2006), who found that distinct neural regions were sensitive to the maximum number (inferior intraparietal sulcus (IPS)) and the complexity (superior IPS and lateral occipital regions) of the items that were held in working memory. At first glance, the proposed dichotomy between number and resolution may appear to conflict with the results of the present experiments, where we have argued that there is an inverse relationship between the number of items that must be stored and the resolution with which they are stored. We suggest that a single account can integrate both conclusions. Specifically, even if there is no relationship between the maximum number of items an individual can hold and the resolution of each memory representation, this does not preclude the possibility that resolution drops as the number of items in memory increases. Thus, our working hypothesis is that there is an average "slot limit" of about 3–4 items, with each slot capable of representing one item regardless of the complexity of that item (Awh et al., 2007). The number of items to be memorized in turn determines the availability of a separate resource that determines mnemonic resolution. The present results suggest that this resource is allocated without regard to the relative complexity of the items to be remembered. Instead, the resources that determine mnemonic resolution are allocated in a discrete slot-based fashion, so that set size and not total information load determines the clarity of each representation in memory.

Acknowledgments

References

Alvarez GA, Cavanagh P. The Capacity of Visual Short-Term Memory Is Set Both by Visual Information Load and by Number of Objects. Psychological Science 2004;15:106–111. [PubMed: 14738517]

Awh E, Barton B, Vogel E. Visual Working Memory Represents a Fixed Number of Items Regardless of Complexity. Psychological Science 2007;18:622–628. [PubMed: 17614871]

Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences 2000;24:87–185. [PubMed: 11515286]

Duncan J, Humphreys GW. Visual search and stimulus similarity. Psychological Review 1989;96:433–458. [PubMed: 2756067]

Eng HY, Chen D, Jiang Y. Visual working memory for simple and complex visual stimuli. Psychonomic Bulletin and Review 2005;12:1127–1133. [PubMed: 16615339]

Ericksen CW, Lappin JS. Selective attention and very short-term recognition memory for nonsense forms. Journal of Experimental Psychology 1967;73:358–364. [PubMed: 6032538]

Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature 1997;390:279–281. [PubMed: 9384378]

Palmer J. Attentional limits on the perception and memory of visual information. Journal of Experimental Psychology: Human Perception and Performance 1990;16:332–350. [PubMed: 2142203]

Pashler H. Familiarity and visual change detection. Perception and Psychophysics 1988;44:369–378. [PubMed: 3226885]

Reicher GM. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology 1969;81:275–280. [PubMed: 5811803]

Rouder JN, Morey RD, Cowan N, Zwilling CE, Morey CC, Pratte MS. An assessment of fixed-capacity models of visual working memory. Proceedings of the National Academy of Sciences 2008;105:5975–5979.

Scolari M, Vogel E, Awh E. Perceptual Expertise Enhances the Resolution But Not the Number of Representations in Working Memory. Psychological Bulletin and Review 2008;15:215–222.

Sperling G. The information available in brief visual presentations. Psychological Monographs: General and Applied 1960;74(11, Whole No. 498)

Todd JJ, Marois R. Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 2004;428:751–753. [PubMed: 15085133]

Vogel EK, Machizawa MG. Neural activity predicts individual differences in visual working memory capacity. Nature 2004;428:748–751. [PubMed: 15085132]

Wilken P, Ma WJ. A detection theory account of change detection. Journal of Vision 2004;4:1120–1135. [PubMed: 15669916]

Xu Y, Chun M. Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 2005;440:90–95.

Zhang W, Luck S. Discrete fixed-resolution representations in visual working memory. Nature 2008;453:233–235. [PubMed: 18385672]
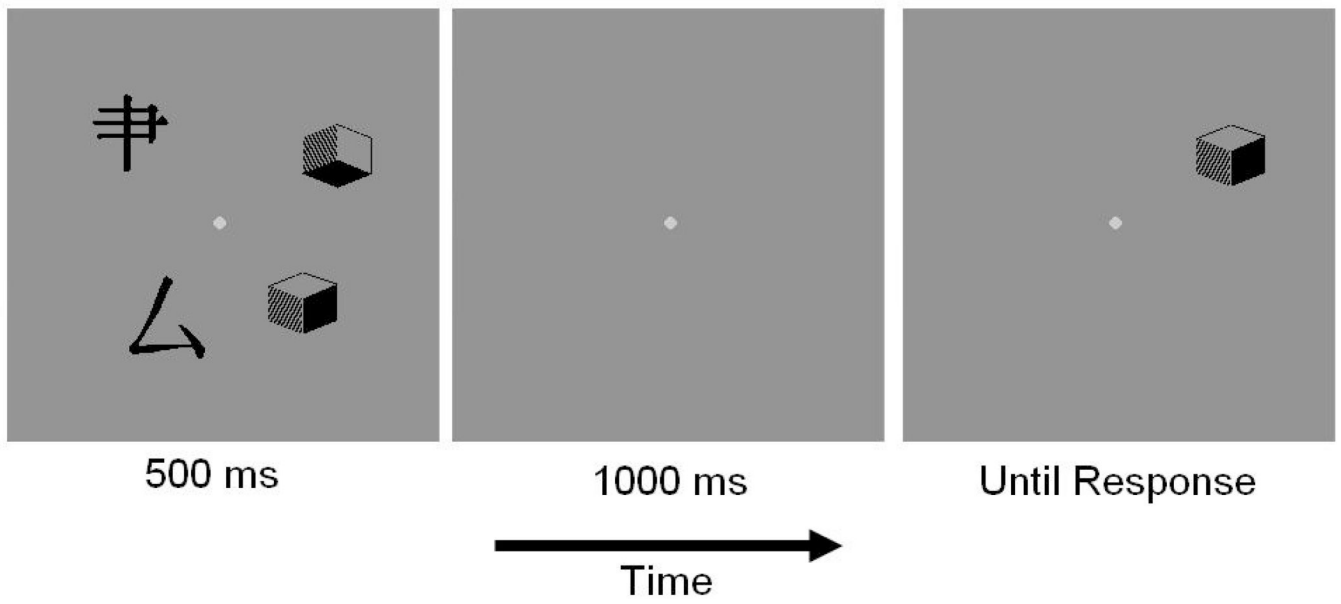
Figure 1.
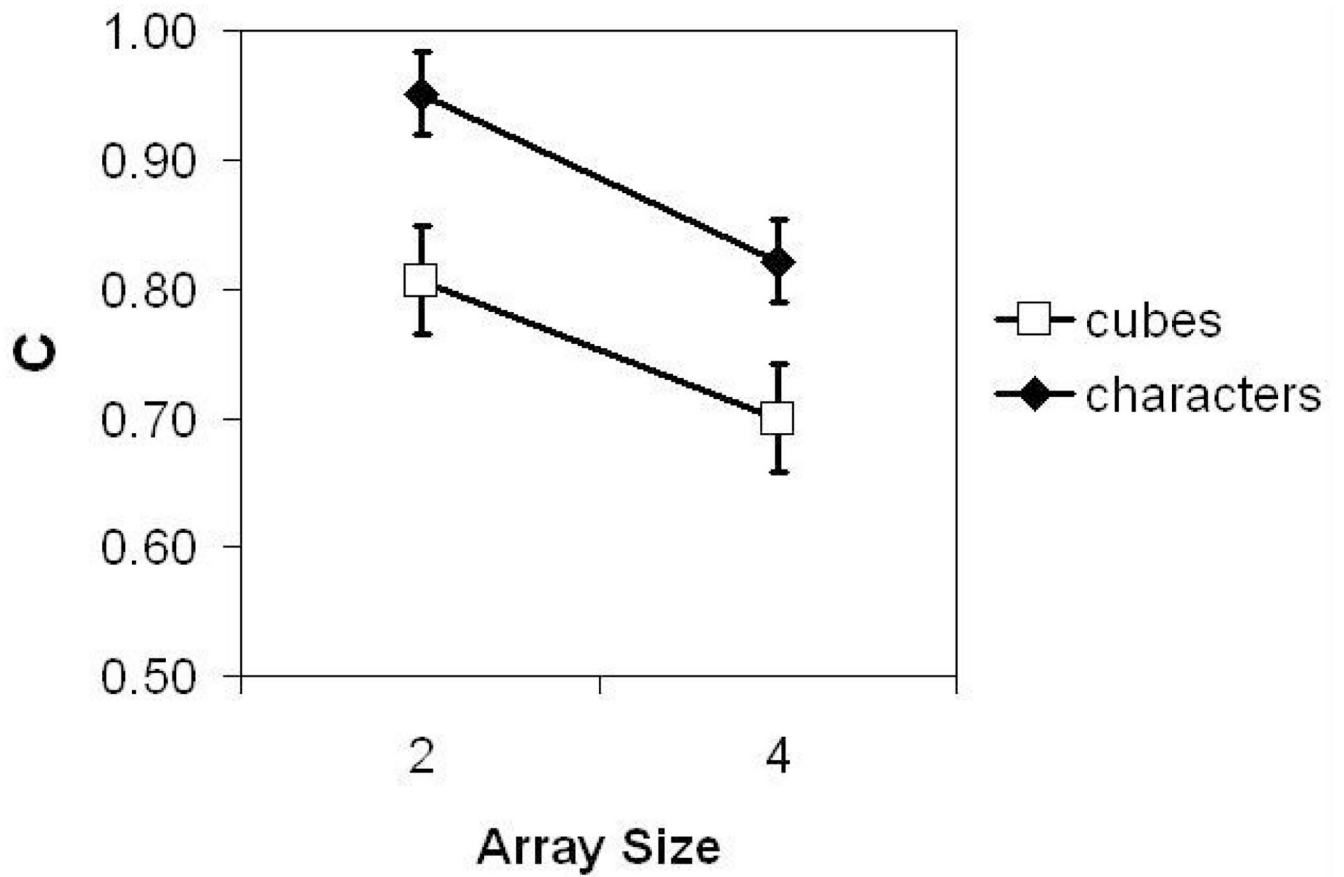The sequence of events in a single trial of Experiment 1.

Figure 2.
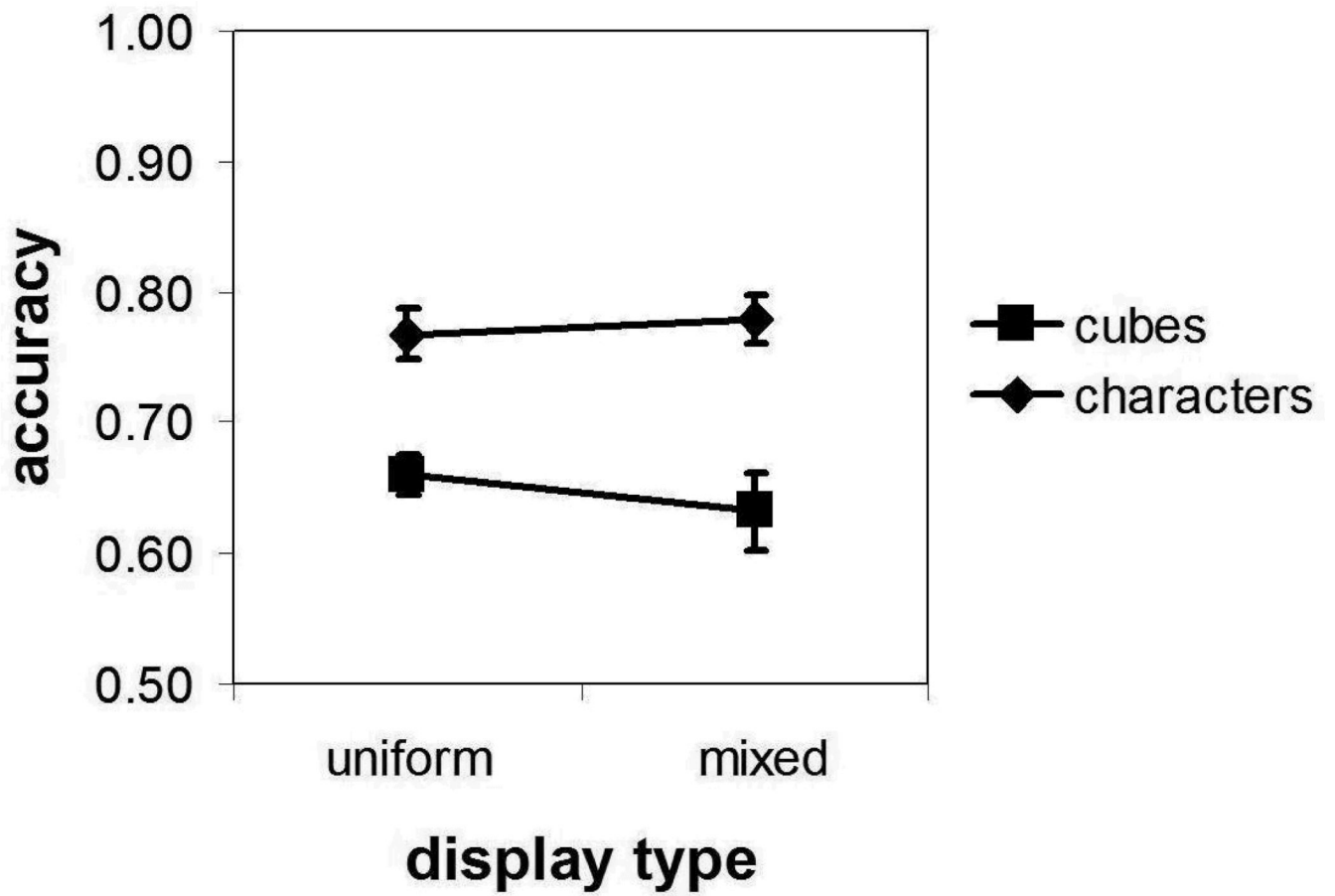Mnemonic resolution, C, in Experiment 1 as a function of object category and set size.

Figure 3.
Change detection accuracy in Experiment 2 as a function of object category and display type.
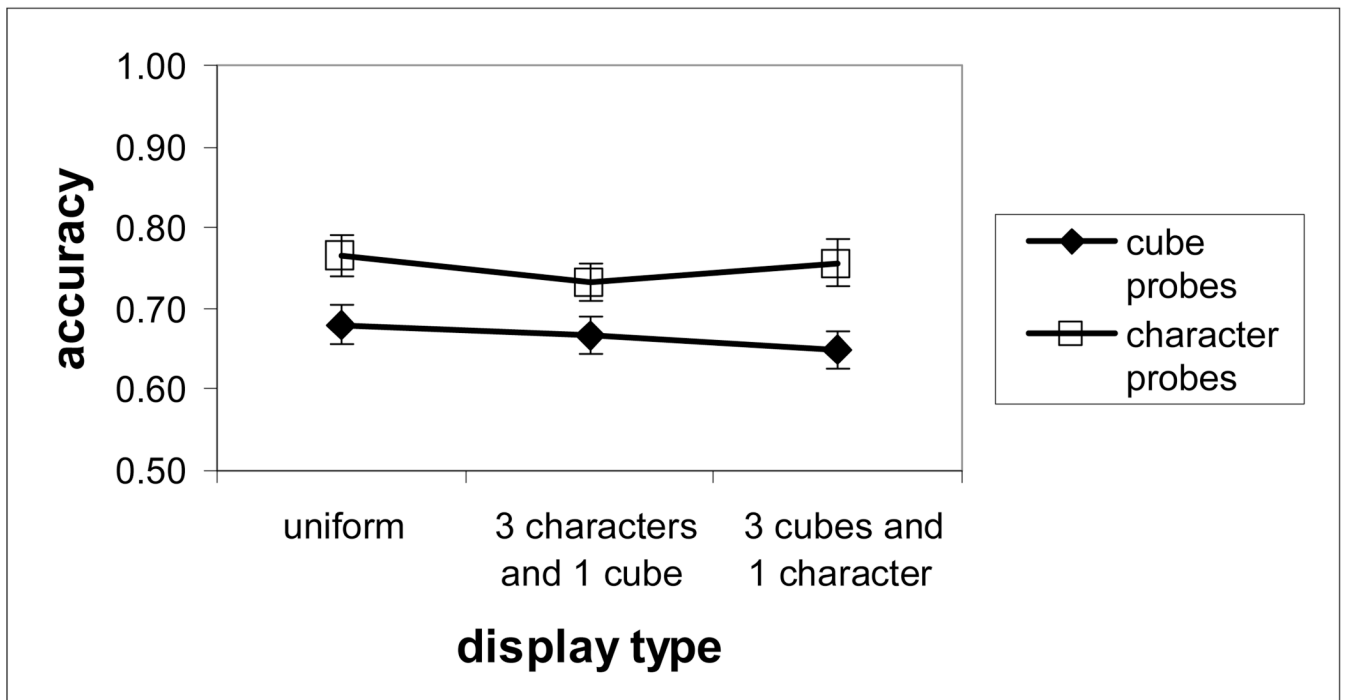
Figure 4.
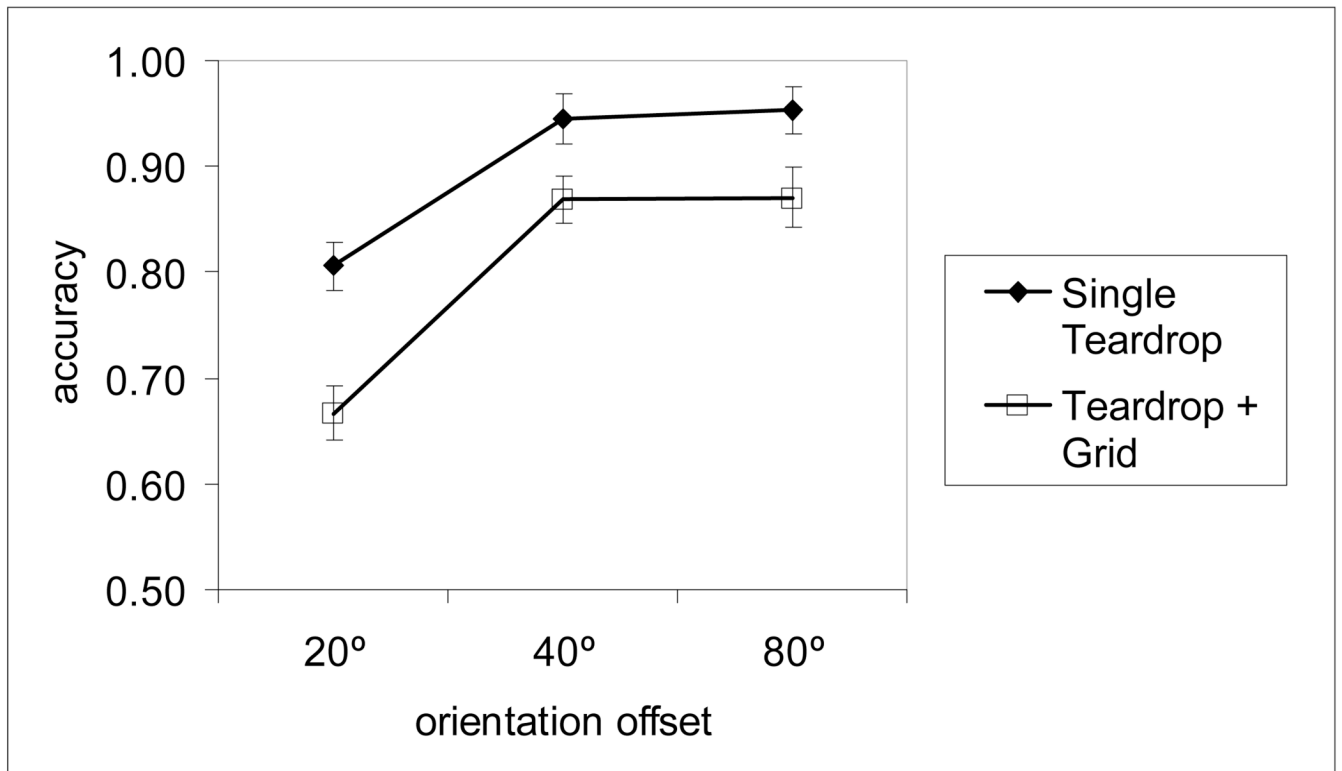Change detection accuracy in Experiment 3 as a function of object category and display type.

Figure 5.
Accuracy with teardrop probes in Experiment 4a as a function of orientation offset and whether or not a grid also had to be stored.
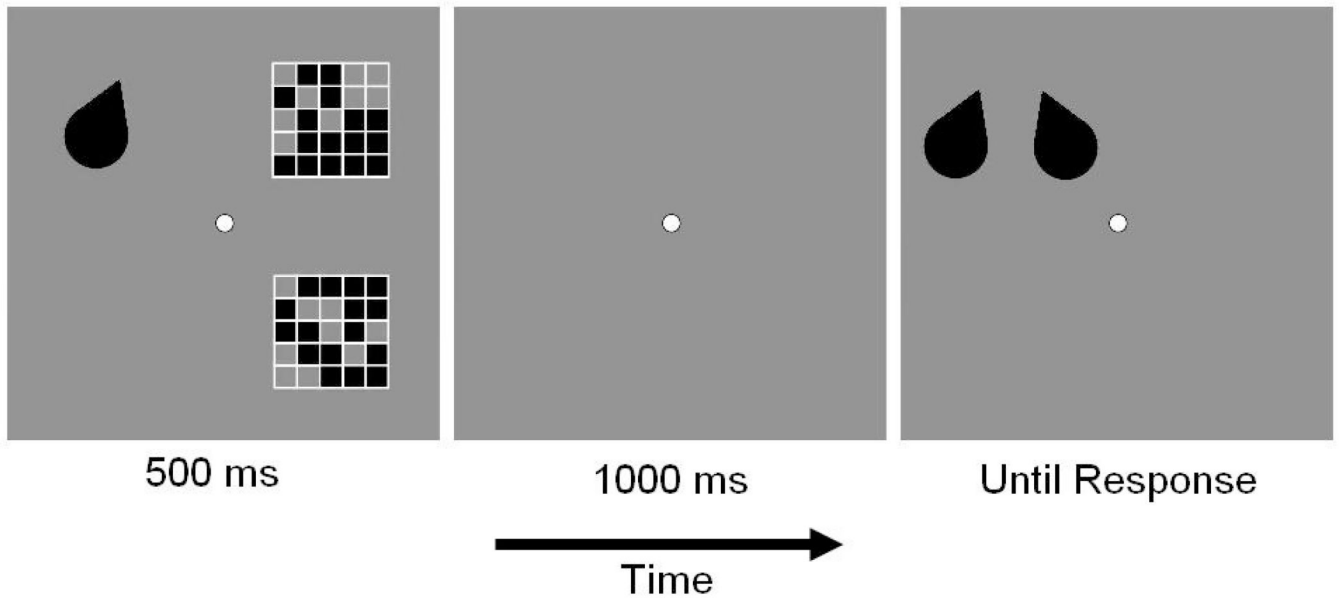
Figure 6.
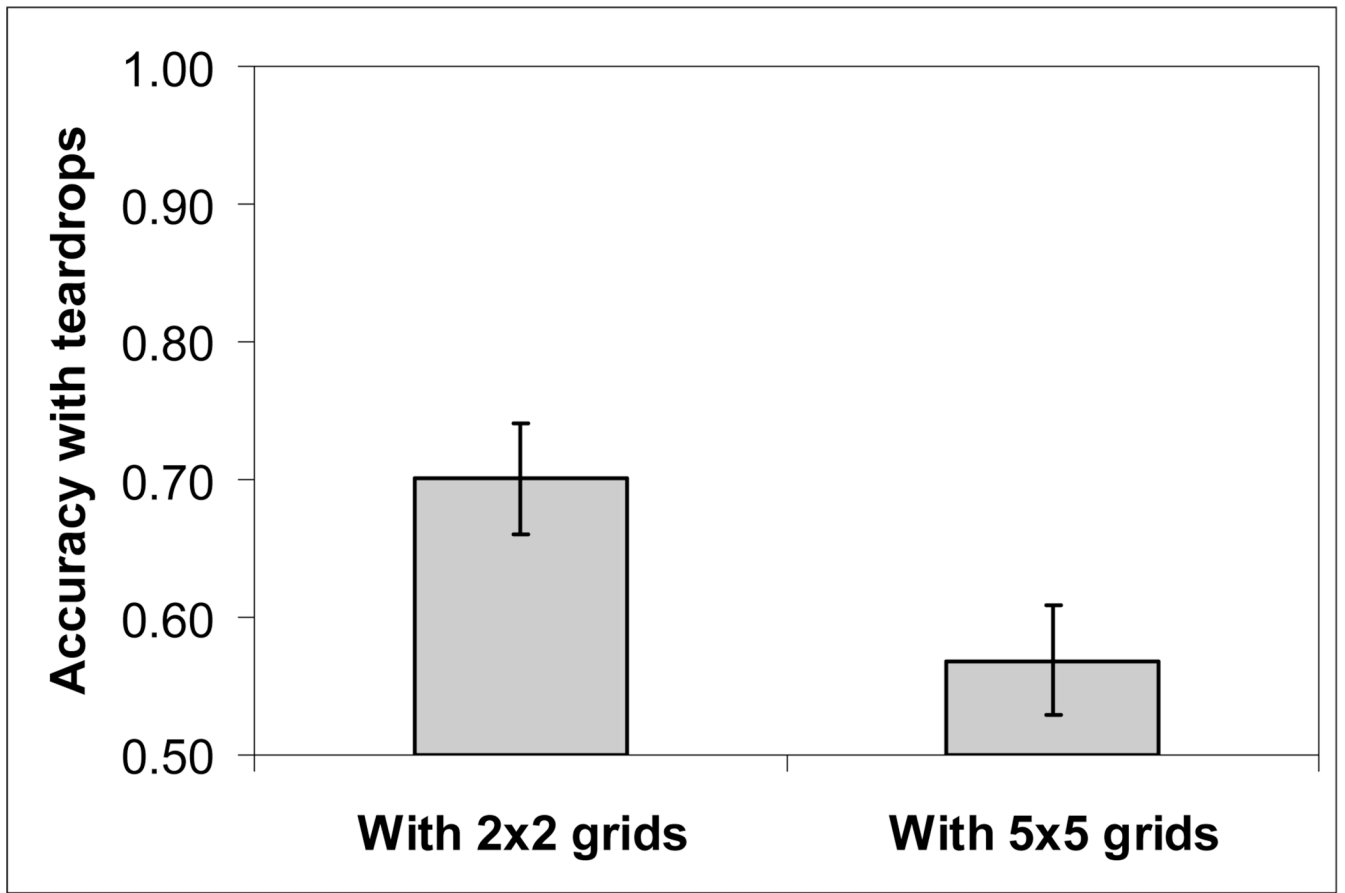The sequence of events in a single trial of Experiment 4b.

Figure 7.
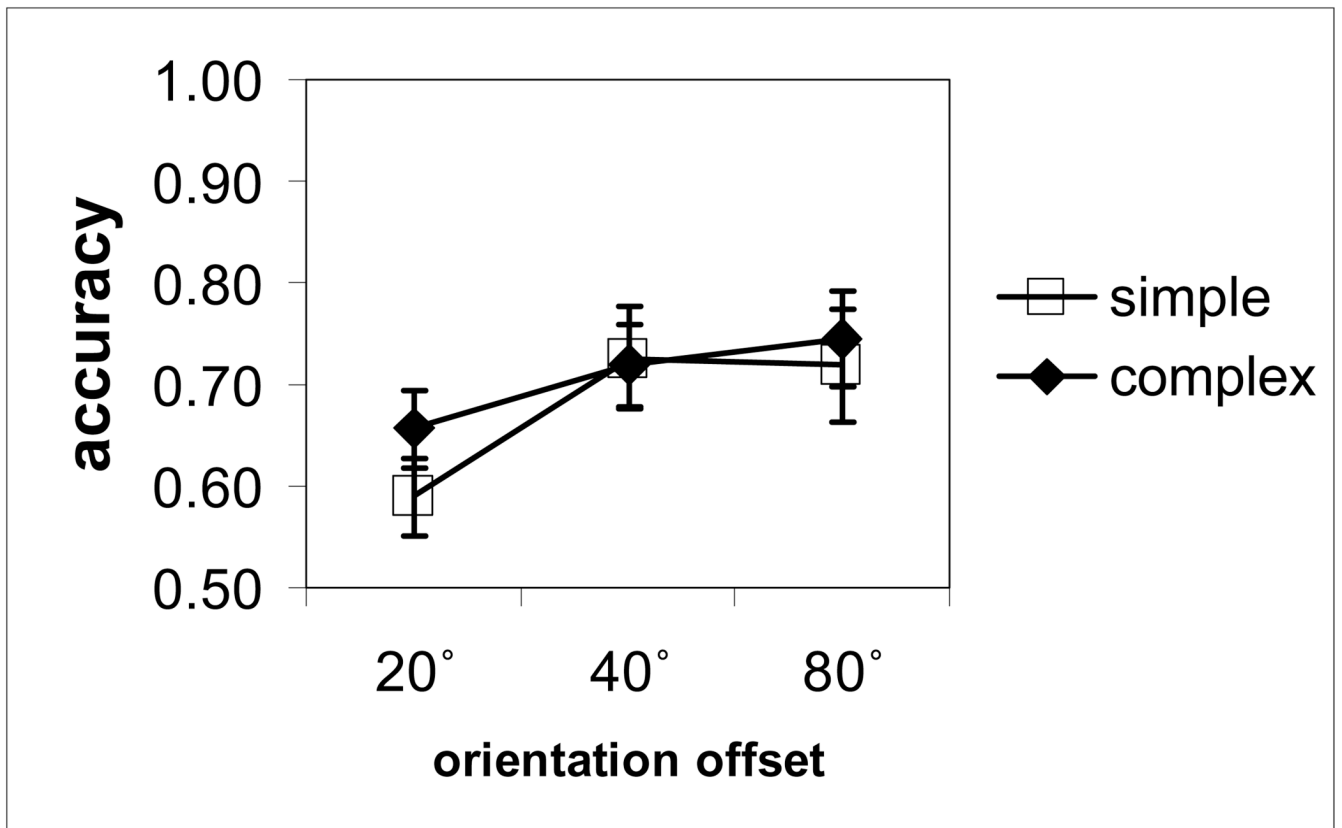Accuracy with simple and complex grids in Experiment 4b.

Figure 8.
Accuracy with teardrop stimuli as a function of angular offset and the complexity of the grids within the same sample display.