Journal of Experimental Psychology:
Learning, Memory, and Cognition
2012, Vol. 38, No. 5, 1432–1438

0278-7393/12/$12.00 DOI: 10.1037/a0027866

RESEARCH REPORT

Grouping and Binding in Visual Short-Term Memory

Philip T. Quinlan
University of York

Dale J. Cohen
University of North Carolina at Wilmington

Findings of 2 experiments are reported that challenge the current understanding of visual short-term memory (VSTM). In both experiments, a single study display, containing 6 colored shapes, was presented briefly and then probed with a single colored shape. At stake is how VSTM retains a record of different objects that share common features: In the 1st experiment, 2 study items sometimes shared a common feature (either a shape or a color). The data revealed a *color sharing effect*, in which memory was much better for items that shared a common color than for items that did not. The 2nd experiment showed that the size of the color sharing effect depended on whether a single pair of items shared a common color or whether 2 pairs of items were so defined—memory for all items improved when 2 color groups were presented. In explaining performance, an account is advanced in which items compete for a fixed number of slots, but then memory recall for any given stored item is prone to error. A critical assumption is that items that share a common color are stored together in a slot as a chunk. The evidence provides further support for the idea that principles of perceptual organization may determine the manner in which items are stored in VSTM.

Keywords: chunking, feature binding, VSTM, slot-based storage, resource-based memory allocation

To operate effectively in a continuously changing environment, one must maintain a record of encountered objects when these move in and out of the immediate visual scene. Theorists have discussed how this is achieved by debating the nature of visual short-term memory (VSTM). Here, we present two experiments designed to study retention of visual information over brief intervals in a bid to examine how information is encoded and then stored in VSTM. Quite unexpectedly, these experiments have produced findings that challenge the main theoretical accounts of VSTM.

Interest in VSTM has increased dramatically since the pioneering work of Luck and Vogel (1997). Starting with their work, an overarching aim in the field has been to try to understand how the different perceptual features of seen objects are coded and maintained over brief intervals in memory. One of the core issues concerns *visual feature binding*—how is a mental record generated and maintained that specifies which features of seen objects go together? The present study addresses this issue by examining how it is that VSTM manages to maintain feature bindings when the constituent features of the to-be-remembered items are shared amongst different objects. That is, assume one display contains two items, namely, a red square and a red triangle, and another display contains a red square and a green triangle. Is memory for the feature bindings in these two kinds of displays equivalent?

There are two general and competing accounts of VSTM that bear on this question. We may refer to these as, respectively, *slot-based accounts* and *resource-limited accounts*. According to slot-based accounts, the primary constraint on VSTM is the number of so-called *slots* that comprise the store (see e.g., Zhang & Luck, 2008). The capacity of VSTM is defined as being equivalent to a fixed number of slots (typically assumed to be three or four; Bays & Husain, 2008; see also Cowan, 2001). Slot-based accounts dovetail with similar ideas about object-based representations. This is because "slots" may be equated with object-based representations such that a slot corresponds to a representation that codes the integrated features of one object (see e.g., Quinlan & Cohen, 2011; Vogel, Woodman, & Luck, 2001). Alternative theories are associated with different assumptions and focus instead on claims that VSTM operates flexibly within constraints set by general resource limitations (cf. Bays & Husain, 2008). According to such views, memory performance is constrained by the demands of coding and maintaining a record of visual features within the limits set by a fixed amount of mental resources. The more information that the system has to deal with then the less likely it is that the memory for the presented material will remain intact.

There is much debate over the veracity of these two different accounts (see Brady, Konkle, & Alvarez, 2011, for a recent review), and it seems that one way to approach the issue is to undertake experiments in which feature duplication across the to-be-remembered items is systematically investigated. By the simplest slot-based account of VSTM, if it is assumed that the slots capture feature conjunctions in bound object representations then there should be little, if any, effect of repeating feature tokens of

This article was published Online First March 26, 2012.

Philip T. Quinlan, Department of Psychology, University of York, York, United Kingdom; Dale J. Cohen, Department of Psychology, University of North Carolina at Wilmington.

Very useful feedback on a preliminary presentation of the data was provided by members of the York Working Memory Group.

Correspondence concerning this article should be addressed to Philip T. Quinlan, Department of Psychology, University of York, Heslington, York, United Kingdom YO10 5DD. E-mail: philip.quinlan@york.ac.uk

the same type across the to-be-remembered items. Memory for a red square should not depend on the number of other red items that are present in the display. In contrast, by the simplest resource-limited account, it seems that memory ought to show some form of benefit if fewer different types of features are presented. So the memory of a red square ought to be better in the presence of other red things simply because the number of different colors that needs to be retained is less than if there were no such repeats. These predictions formed the basis for the first experiment.

Of course, there are many variations on how VSTM may be limited (cf. Wheeler & Treisman, 2002). However, a thorough discussion of these alternatives can be avoided because the actual data described here reveal effects that were quite unexpected and do not fit easily within any of the current theories of VSTM.

Experiment 1

We used a version of the *change detection paradigm* (see Rouder, Morey, Morey, & Cowan, 2011) known as *single-probed recognition*. In this case, and on every trial, an initial study display that contained six colored shapes was presented briefly. This was removed and then, following a brief retention interval, a second probe display that contained a single colored shape (i.e., the probe) was presented. The participant had to make a present/absent judgment as to whether the probe had been present in the study display. On positive trials, the probe was a repeat of a colored shape in the study display. On negative trials, the probe comprised a recombination of a color and a shape from the study display. In this way, success at the task was predicated on recognizing particular color–shape bindings from the study display. Participants were tested on their ability to remember exactly which colors went with which shapes in the study display.

Method

Participants. Originally 34 participants were tested, but for four individuals the level of accuracy across all conditions was very poor (i.e., less than 55% correct), and their data were discarded. Therefore, the final sample consisted of 30 participants (mean age = 20.1 years, *SD* = 1.6; five were male; three were left handed). They were enlisted from the undergraduate participant panel at the University of York. All reported having normal or corrected-to-normal vision. None reported having any color vision problems.

Materials and design. The stimuli were sampled from eight solid shapes taken from the SPSS Marker Set Font (see Figure 1A), and each shape could be presented in one of eight colors (see Quinlan & Cohen, 2011).

On each trial in the experiment a study display was presented prior to a probe display. Each study display contained six colored shapes (each 1.0° × 1.0°) equally spaced around a virtual circle (radius 1.5°) centered at fixation. Prior to each trial, the individual shapes and colors were sampled at random within the particular constraints that were defined by the different conditions. In the probe display the colored shape was presented at the central fixation.

The five key conditions were a baseline condition, two repeated color conditions, and two repeated shape conditions. For each of these five conditions, probe present and probe absent cases were

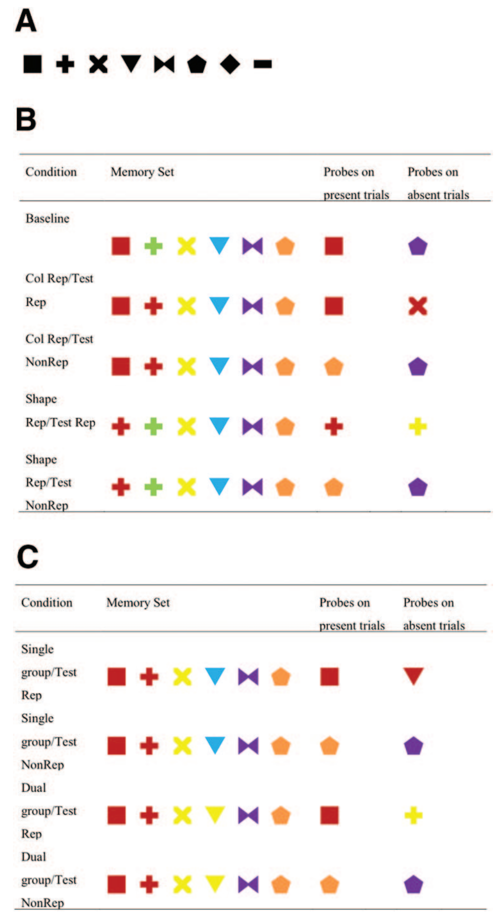

Figure 1. Stimuli and schematic representations of the various trial types in the experiments. A: The eight basic shapes used in the experiments. B: Schematic representation of the 10 conditions of interest in Experiment 1. C: Schematic representation of the eight conditions of interest in Experiment 2.

generated (see Figure 1B). In the baseline condition, six different shapes were chosen, and each was associated with a different color. In the color repeated/test repeated condition, two items shared a common color in the study display, and the probe corresponded to one of the items with the repeated color. In the color repeated/test nonrepeated condition, the study display was identical to the color repeated/test repeated condition, but the probe corresponded to one of the items in which the color was not repeated. In the shape repeated/test repeated condition, two items shared a common shape in the study display, and the probe corresponded to one of items with the repeated shape. In the shape repeated /test nonrepeated condition, the study display was identical to the shape repeated/test repeated condition, but the probe corresponded to one of the items in which the shape was not repeated. For each of the repeated color and repeated shape conditions, two probe absent cases were defined: one in which the repeated feature was present in the probe and one in which only unique features were present.

Present and absent cases were paired up as shown in Figure 1B, giving 10 basic conditions. Following an initial block of 10 practice trials, there were four blocks of experimental trials. The order of trials within each block was randomized on a participant-by-participant basis. Each of the 10 basic conditions was tested via 40 experimental trials.

Procedure. Each participant was tested individually in a quiet testing room. Participants sat at a table facing a 17-in. computer monitor placed on a plinth so that the center of the screen was at eye level. A keyboard was placed in front of the screen, and responses—a "1" for probe present and a "2" for probe absent—were made via the keyboard. Viewing distance from the screen was approximately 60 cm. The experiment was controlled by a PC running E-Prime (Schneider, Eschman, & Zuccolotto, 2002).

Participants were instructed as to the nature of the task and that accuracy and not speed of response was the primary measure of interest. They were allowed to complete the first block of practice trials and then asked to sign a consent form. At the start of every trial a central black dot acted as the fixation point and was presented for 500 ms. (The screen background was "silver," a defined Windows color, throughout.) This was replaced immediately by the study display for 250 ms. A blank screen was then presented for 900 ms, and finally the probe display was presented until response. Once a response was detected, corresponding feedback—the word "Correct" or "Error"—was presented for 500 ms, and this led immediately on to the fixation dot for the next trial. A rest break was scheduled at the end of each block, and the participant initiated the next block with the press of the mouse.

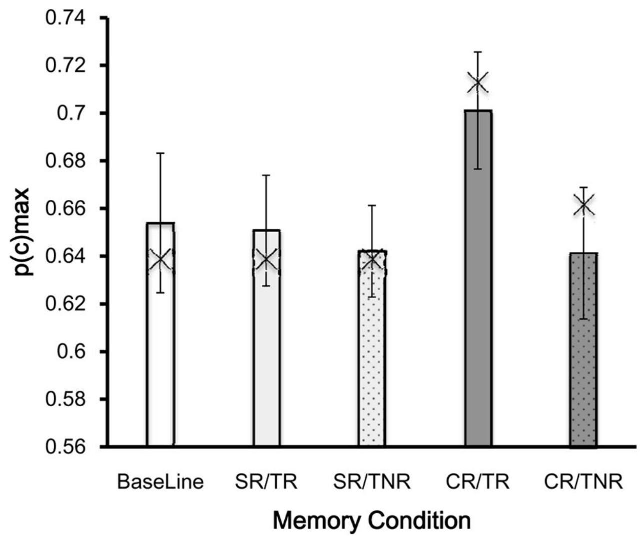

Figure 2. Graphical representation of the memory performance scores (expressed in terms of $p(c)_{max}$) for the various memory conditions in Experiment 1. Error bars reflect 95% confidence intervals as specified by Morey (2008). The "X" symbols indicate the corresponding model predictions for the various conditions (see General Discussion). SR/TR = shape repeated/test repeated; SR/TNR = shape repeated/test nonrepeated; CR/TR = color repeated/test repeated; CR/TNR = color repeated/test nonrepeated.

Results and Discussion

The overall average level of accuracy across all participants across all conditions was 65%. In line with the advice provided by Pastore, Crawley, Berens, and Skelly (2003), $p(c)_{max}$ scores were derived for each participant in each of the five key conditions. The term $p(c)_{max}$ refers to the unbiased proportion correct that may be computed from d' and is linearly related to d' (after Macmillan & Creelman, 1991). As such, $p(c)_{max}$ is a sensitivity measure that is independent of response bias. A graphical illustration of average performance in each of the key conditions is shown in Figure 2. As can be seen from the figure, item recall was particularly good when a color was repeated in the study display and the probe shared the repeated color. Hit and false alarm rates for the conditions of interest are displayed in Table 1.

Initially, $p(c)_{max}$ scores for the repeated feature trials were entered into a 2 × 2 repeated-measures analysis of variance (ANOVA) in which repeated feature (color, shape) and probed item (repeated features, unique features) were entered as fixed factors. This analysis revealed statistically reliable main effects of repeated feature, $F(1, 29) = 8.98$, $MSE = 0.002$, $p < .01$, and probed item, $F(1, 29) = 10.76$, $MSE = 0.003$, $p < .01$, together with a statistically significant Repeated Feature × Probed Item interaction, $F(1, 29) = 6.47$, $MSE = 0.003$, $p = .017$. In order to explore this pattern of performance further, a Tukey's honestly significant difference (HSD) test was carried out on the corresponding cell means. This test revealed that performance in the repeated color/test repeated condition was overall the best ($p < .05$ for all comparisons). No other pairwise comparisons reached statistical significance (all $ps > .05$).

An additional Dunnet's test was carried out in order to compare performance in the baseline condition (in which only unique colors and shapes were presented) with performance in the individual repeated feature conditions. This test revealed that the only condition in which performance differed from baseline was the repeated color/test repeated condition ($p < .05$; for all other pairwise comparisons, $ps > .05$). Therefore, relative to the baseline condition, participants' memory performance was best in the repeated color/test repeated condition.

In sum, the findings are both clear-cut and unpredicted. Performance was overall best in the repeated color/test repeated condition. If we focus on the displays in which a color was repeated, memory for particular color–shape bindings was better when the corresponding items shared a repeated color than when they did not. This particular finding is referred to as the *color sharing effect*. The size of this color sharing effect was determined via a within-participant contrast (Keppel & Wickens, 2004, pp. 357–359) and accords with Cohen's (1988) definition of a "large" effect ($d = 1.3$).

The color sharing effect does not naturally follow from either the simple slot-based account or the resource-limited account sketched previously. The slot-based account provides no explanation of the color sharing effect because the number of the to-be-remembered items was the same in all conditions. This finding is also awkward for the simple resource-limited account. It predicts increased memory performance for all elements in a given display when fewer feature types are to be remembered. There should be no selective improvement only for those items that share a common color. Clearly the properties of VSTM, as revealed by the current data, are quite unexpected given the simple slot-based and resource-limited accounts as set out previously.

The results also revealed a clear difference in performance across the repeated color and repeated shape displays. Whereas

Table 1
Hit and False Alarm Rates for the Conditions of Interest in Experiment 1

Condition	p(Hits)	p(FA)
Baseline	.60	.30
Shape rep/Test rep	.61	.32
Shape rep/Test nonrep	.59	.32
Color rep/Test rep	.76	.38
Col rep/Test nonrep	.59	.31

Note. Data are expressed as proportions. Rep = repeated; nonrep = nonrepeated; FA = false alarm.

there was a benefit in the cases where the bindings shared a common color, there was no such benefit in the cases where the bindings shared a common shape. Visual inspection of the displays gave rise to the impression that shared color was a much stronger cue to grouping than shared shape. Indeed in a related study in which the same combinations of colors and shapes were used (i.e., Quinlan & Cohen, 2011), it was found that the colors were more discriminable from one another (as given by measures of d') than were the shapes. In other words, the colors were more distinguishable than were the shapes. On these grounds, we argue that here color acted as a more salient grouping cue than did shape. An implication of this is that the color sharing effect provides further evidence that VSTM performance can reflect sensitivities to perceptual processes of item grouping (Jiang, Chun, & Olson, 2004; Woodman, Vecera, & Luck, 2003). The data show that if the display contained a distinctive group of items that shared a common color then the memory for those items was better than it was for items that were not so grouped.

This is not the first time that a special role for color in VSTM has been documented. The nature of the current color sharing effect bears some similarity with the findings reported by Lin and Luck (2008). In their final experiment, and on each trial, participants were presented with sequence of study displays, each of which contained a colored square patch. Each sequence contained three such study displays, and the position of the patch changed across the displays. Following the study displays, a single probe item was presented, and participants had to judge whether the probe matched the color of the study patch at the same position.

Across each sequence, two of the color patches were of a similar hue and the third was of a distinctive hue. The central finding was that memory was better for items of a similar hue than it was for the item presented in a distinctive hue. The similarity of this finding with the present color sharing effect is notable. Whereas Lin and Luck (2008) have shown that memory for particular colors is enhanced if similar colors are being maintained in VSTM, we have shown that memory for particular color–shape bindings is enhanced if the corresponding items in VSTM share the identical color. It seems that both cases point to the potency of grouping by color within VSTM (see also Brady & Alvarez, 2011).

As neither the simple slot-based nor the resource-limited theory of VSTM provides a ready explanation of the color sharing effect, a second experiment was undertaken that allowed us to assess more generally the importance of grouping processes within VSTM. In line with the Gestalt principles of grouping by similarity (see Quinlan & Dyson, 2008, Chapter 5), we claim that the items that shared a common color were treated as a group and were stored and maintained as such in VSTM. Moreover, the data suggest that the bindings within such groups are maintained better in VSTM than are the bindings associated with uniquely colored items. Implications of this suggestion were tested in a second experiment.

Experiment 2

The rationale for the second experiment was partially inspired by some findings reported by Kahneman and Henik (1977) in the context of experiments on free recall of grouped items. In one of their experiments each study display contained a single row of digits in which the items were grouped according to spatial proximity, that is, “1234 56” or “123 456.” Recall of digits from the first (left-most) group was consistently better than recall of items from the second. In order to explain this sort of result, Kahneman and Henik developed a group-processing model in which the idea of the allocation of a limited capacity processing resource was discussed. According to this account, items are initially grouped together according to some principle of perceptual organization, and, in cases where more than one group is formed, a queue is set up in order for the groups to access a putative, finite pool of processing resources. The first group of items in the queue immediately draws on the pool of resources, and whatever remains after this is then made available to the next group in the queue. This sort of account provides something of an explanation of the color sharing effect on the grounds that the items grouped by a common color accrue more resources than the items in the display that do not share a common color.

The question now was, therefore, whether the color sharing effect would vary according to the presence of more than one salient group of items. For instance, if fewer resources are distributed to a second group of items than the first (as shown in the results of Kahneman & Henik, 1977) then, on average, performance with the repeated color items would be worse in displays containing two shared colors than those containing a single shared color. As the size of the color sharing effect is, in part, determined by performance with the items with distinctive colors then, depending on the manner in which resources are allocated to these items, different predictions follow. Nonetheless, they all predict that, on average, performance with the uniquely colored items should also suffer if the allocation of resources accords with the sort of group-processing account advocated by Kahneman and Henik (1977).

It is possible to consider a variety of resource-allocation models of the kind discussed by Kahneman and Henik (1977), all of which predict generally poorer performance with displays containing dual groups than with those containing one such group. Quite different predictions arise if it is assumed that the allocation of resources is determined simply by the number of features that need to be coded. For instance, if displays containing fewer features place fewer demands on resources than displays containing more features, then performance overall will be better when the displays contain two color groups than a single color group. In contrast to all such resource-allocation accounts, the original slot-based account predicts no difference in performance with the various kinds of displays.

Therefore, in Experiment 2 two new conditions were tested, and, in both, two colors were repeated within a study display such that two shapes shared one color and two different shapes shared a different color. Performance with these displays was now tested alongside the original repeated color conditions (see Figure 1C for a schematic breakdown of the conditions of interest).

Method

In nearly all respects, the method was as before. Data from a new sample of 30 participants from the undergraduate participant panel at the University of York (mean age = 22.1 years, *SD* = 3.8; four were male; all were right handed) are reported. All participants reported having normal or corrected-to-normal vision. None reported having any color vision problems. Six individuals were replaced due to poor levels of accuracy at the task (i.e., their overall levels of accuracy were less than 55%).

Design. The repeated color/test repeated and the repeated color/test nonrepeated conditions were included as before. These are now referred to as the single group conditions. In addition, dual group conditions were also generated (see Figure 1C). In these cases, each study display consisted of two items that shared one color, two items that shared a different color, and two items that each had a unique color (i.e., six items and four colors). Probe present and probe absent trials were generated for each of these four key conditions as shown in Figure 1C, giving eight basic conditions in total. Four blocks of experimental trials followed a block of eight practice trials. In total there were 46 experimental trials for each of the eight key conditions.

Results and Discussion

The average level of accuracy over all participants over all conditions was 63%. As in Experiment 1, performance was expressed in terms of $p(c)_{max}$ scores, and a graphical summary of these scores is shown in Figure 3. As the figure clearly shows, the same color sharing effect found in Experiment 1 was found here. However, no such effect was present in the data for the dual group conditions. Hit and false alarm rates for the conditions of interest are displayed in Table 2.

The corresponding $p(c)_{max}$ scores for the four key conditions were entered into a 2 × 2 repeated-measures ANOVA in which grouping (single group vs. dual group) and probed item (repeated features, unique features) were entered as fixed factors. This analysis revealed statistically reliable main effects of grouping, $F(1, 29) = 7.24$, $MSE = 0.003$, $p = .012$, and probed item, $F(1, 29) = 20.09$, $MSE = 0.002$, $p < .0001$. In addition, the Grouping × Probed Item interaction, $F(1, 29) = 5.36$, $MSE = 0.002$, $p < .05$, also reached statistical significance.

A Tukey's HSD test revealed the detailed nature of this interaction. Performance was overall worst in the color repeated/test nonrepeated condition (all ps < .05). No other pairwise comparison reached statistical significance (all ps > .05). Nonetheless, the original color sharing effect as reported in Experiment 1 was present in the data for the single group conditions. As before, the size of this color sharing effect was determined via a within-participant contrast and accords with Cohen's (1988) definition of a "medium" to "large" effect ($d = 0.7$).

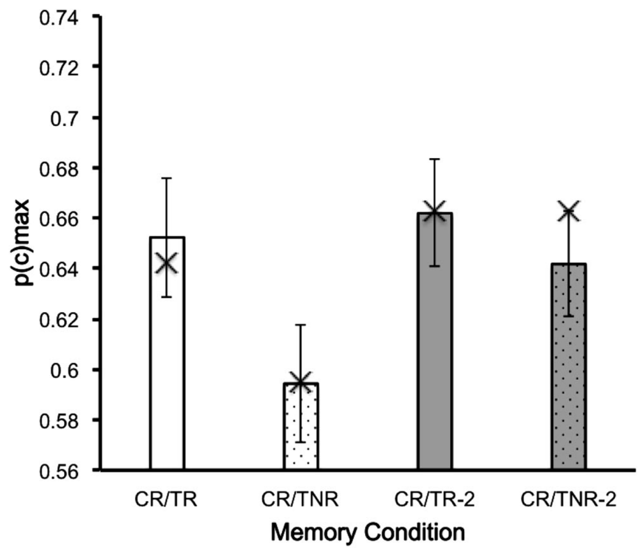

Figure 3. Graphical representation of the memory performance scores (expressed in terms of $p(c)_{max}$) for the various memory conditions in Experiment 2. Error bars reflect 95% confidence intervals as specified by Morey (2008). The "X" symbols indicate the corresponding model predictions for the various conditions (see General Discussion). CR/TR = color repeated/test repeated; CR/TNR = color repeated/test nonrepeated; "-2" = dual group condition.

In sum, the data revealed that there was a statistically robust color sharing effect in the data for the single group conditions—participants were more accurate when probed by an item that shared a color with another item in the study display than when the probed item was associated with a unique color. However, there was no such color sharing effect in the data for the dual group conditions. In these conditions participants were as accurate in their reports of items that shared a color as they were in their reports of items that had distinctive colors. Importantly, the presence of two colored groups of items increased participants' performance with the uniquely colored items when assessed relative to performance with the same items in the single group condition.

General Discussion

In Experiment 1 the central finding was that participants' memory was best when the probed item shared a common color with another study item than when it did not. On the surface, this finding appears to support the resource-allocation account described by Kahneman and Henik (1977) in their group-processing model. According to this account, if a display contains a salient group of items then this group will accrue proportionally more resources than the other ungrouped items in the display. However, this account predicts not only a performance benefit for the grouped items but also an associated cost in performance with the ungrouped items. The data from Experiment 1, however, do not accord well with this account. Although there was an overall benefit in performance for grouped items, there was no discernible cost in performance with the other items in the display. Performance with the uniquely colored items (as gauged by performance

Table 2
Hit and False Alarm Rates for the Conditions of Interest in Experiment 2

Condition	p(Hits)	p(FA)
Single group/Test rep	.68	.39
Single group/Test nonrep	.53	.34
Dual group/Test rep	.70	.38
Dual group/Test nonrep	.53	.26

Note. Data are expressed as proportions. Rep = repeated; nonrep = nonrepeated; FA = false alarm.

in the repeated color/test nonrepeated condition) was no worse than performance at baseline (see Figure 2).

The proposition that grouped items accrue more resources than ungrouped items is further challenged by the data reported in Experiment 2. Although the original color sharing effect was present again when two items in the displays shared a common color, no such color sharing effect was present when two color groups were present in the displays. In the dual group condition, there was a clear benefit and no cost in performance with the ungrouped items (see Figure 3).

It is also a struggle for current characterizations of VSTM as comprising a fixed number of slots (Luck & Vogel, 1997) to accommodate the present findings. For instance, according to the *constant-capacity hypothesis* (Cowan, Chen, & Rouder, 2004), there is a fixed number of slots, and each slot captures a single chunk of material to be remembered. The idea is that although the amount of information contained within a chunk may vary, the number of remembered chunks remains constant. The amount of information stored in any one slot is independent of that stored in any other slot.

An alternative slot-based account has been discussed by Alvarez and Cavanagh (2004). When they varied object complexity across a range of change detection tasks, they found that the estimates of memory capacity scaled inversely with object complexity. They therefore concluded that "more capacity must be allocated to more complex objects" and, furthermore, that "there is a trade-off between complexity of the objects and the total number of objects that can be stored in memory" (p. 109).

Neither of these slot-based accounts can easily accommodate the present data. The data from Experiment 2 show that there was a general improvement in memory for all items when two color groups were present in the displays. This particular grouping effect is not predicted by the constant-capacity hypothesis. Nor is it predicted by the trade-off account discussed by Alvarez and Cavanagh (2004).

Modeling the Current Data Sets

We conclude that the data are problematic for several of the most popular current models of VSTM and therefore that an alternative account is needed. Various types of slot-based and resource-based explanations were considered, but part of the problem is that these have been fleshed out in terms of memory for simple features. A distinctive aspect of the current work is that the experiments concern memory for feature bindings. On these grounds, modeling how feature bindings are stored and maintained in VSTM is key.

Evidence from our previous work leads us to the conclusion that object-based representations are stored and maintained in VSTM (Quinlan & Cohen, 2011). On these grounds, we accept a slot-based account in which each slot captures an object or proto-object. We propose that, in contrast with the previously mentioned models, a slot can accommodate either an individual object or a perceptual group. That is, perceptual encoding ensures that some form of chunking takes place based on principles of perceptual organization. As such, each perceptual group (i.e., a chunk) is taken to correspond to a distal object and may be stored in a slot.

This is not the first time that intimate connections have been drawn between principles of perceptual organization and the structure of VSTM (see Jiang et al., 2004; Woodman et al., 2003), but here we go further and suggest that aside from grouping by proximity (Woodman et al., 2003), grouping by color is a powerful grouping principle. Our first assumption is that items that share a common color are grouped together and furthermore that such grouped items are coded and stored as a chunk in memory. Our second assumption is that VSTM is characterized as comprising a fixed, small number of slots (after Luck & Vogel, 1997). In our ensuing account we assume four such slots and that each slot may capture one distinctive item or, equally, chunked items grouped by a common color. In agreement with the constant-capacity hypothesis (Cowan et al., 2004), we accept that once items have been encoded as chunks, the storage of information in one slot is independent of the information in any other slot.

A further assumption we take from the work of Kahneman and Henik (1977). They pointed out that whatever code is generated during memory encoding it then must gain access "to some storage system that has a capacity of about four items"; that is, there is "competition for storage capacity" (p. 323). We accept the idea that following encoding, inputs compete for the slots. We take it that there is a slot-selection stage of processing during which the probability that a particular input is selected for storage is determined. We assume that all inputs have an equal likelihood of being selected for a slot with the proviso that both uniquely colored items and items grouped by a common color constitute an input. In the event that two items are grouped together, both items will be stored in a single slot if either item is selected. This assumption has the practical effect of doubling the probability of selection for grouped items. The selection process occurs without replacement until all four slots are filled. Therefore, if there are only four groups/items, all four groups/items will be selected for storage regardless of their original selection probabilities. To determine the average probability of selection (termed p) for Experiments 1 and 2, we have run Monte Carlo simulations of the slot-selection stage of processing as just described.

Once an input is stored, it may be recalled. The probability of correct recall is some monotonic function of the selection probability. We have modeled this process via a standard exponential function, namely,

$$0.5 \times e^{\{k \times [1-(p-0.5)]\}} + .05, \tag{1}$$

where k is the fallibility rate (the free parameter in our model) and p is the probability of selection. This is a negatively decelerating function that flattens out at 0.5, which defines chance in our experiments. According to this function, the more likely it is that an input is selected for storage (e.g., 90% storage probability), the larger the likelihood is that some noise will enter the system and an error will occur. This is simply a probability function and is not influenced by the content of the input.

In modeling the data for Experiment 1 (see Figure 2), it was assumed that whereas items that shared a common color were treated as a single input, items that shared a common shape were treated as separate inputs. The model fits for Experiment 2 are shown in Figure 3. According to this account, the relatively good performance with the uniquely colored items in the dual group condition is due to the fact that the number of encoded inputs is equal to the number of available slots. The data for Experiments 1 and 2 were modeled separately, and different estimates for k were derived from the model fits. For Experiment 1, $k = -1.5$ ($SD = 0.05$), and for Experiment 2, $k = -2.3$ ($SD = 0.08$). Experiment 2 had a smaller k because the average performance was lower relative to that in Experiment 1. The model fit both data sets well, with the model accounting for 62% of the variance in Experiment 1 and 81% of the variance in Experiment 2.

In sum, we have been able to account for the data with a simple fixed slot-based account built upon two factors. The primary factor is slot competition: Inputs compete for a fixed, small number of slots. An important proviso is that an input may correspond to a set of items grouped by common color. The secondary factor is memory fallibility. We have modeled a process whereby the more likely it is that an input enters the store, the more likely it is that memory for that input will fail. There is one free parameter in the model, namely, k, the fallibility rate. Our attempts at modeling the data were unsuccessful without this parameter.

Conclusions

The data from two single-probe recognition experiments have shown that when items share a common color, the memory for the corresponding color–shape bindings is much better than it is for items that possess unique colors. This color sharing effect has been explained in terms of how principles of perceptual organization govern the chunking of items that are then stored in VSTM. We assume that VSTM comprises four fixed slots such that any given slot captures either a uniquely colored item or a group of items that share a common color. We have modeled performance by assuming that a critical stage in processing involves inputs competing for slots. The probability of successful access to a slot then determines the degree to which the corresponding input is remembered.

We have shown how a simple single-factor slot-based model successfully accounts for the data. It will be interesting to see if a more adequate account of VSTM performance can be developed when more sophisticated resource-allocation models are put forward.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set by both visual information load and by number of objects. *Psychological Science, 15,* 106–111. doi:10.1111/j.0963-7214.2004.01502006.x

Bays, P. M., & Husain, M. (2008, August 8). Dynamic shifts of limited working memory resources in human vision. *Science, 321,* 851–854. doi:10.1126/science.1158023

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. doi: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,* 4–34. doi:10.1167/11.5.4

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science, 15,* 634–640. doi:10.1111/j.0956-7976.2004.00732.x

Jiang, Y., Chun, M. M., & Olson, I. R. (2004). Perceptual grouping in change detection. *Perception & Psychophysics, 66,* 446–453. doi:10.3758/BF03194892

Kahneman, D., & Henik, A. (1977). Effects of visual grouping on immediate recall and selective attention. In S. Dornič (Ed.), *Attention and Performance VI* (pp. 307–332). Hillsdale, NJ: Erlbaum.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Lin, P.-H., & Luck S. J. (2008). The influence of similarity on visual working memory representations. *Visual Cognition, 17,* 356–372. doi:10.1080/13506280701766313

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

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4,* 61–64.

Pastore, R. E., Crawley, E. J., Berens, M. S., & Skelly, M. A. (2003). "Nonparametric" A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10,* 556–569. doi:10.3758/BF03196517

Quinlan, P. T., & Cohen, D. J. (2011). Object-based representations govern both the storage of information in visual short-term memory and the retrieval of information from it. *Psychonomic Bulletin & Review, 18,* 316–323. doi:10.3758/s13423-011-0064-2

Quinlan, P. T., & Dyson, B. J. (2008). *Cognitive psychology.* Harlow, England: Pearson.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review, 18,* 324–330. doi:10.3758/s13423-011-0055-3

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114. doi:10.1037/0096-1523.27.1.92

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

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10,* 80–87. doi:10.3758/BF03196470

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

Received July 20, 2011
Revision received February 6, 2012
Accepted February 7, 2012 ■