Running head: PRE-CATEGORICAL VSTM

The pre-categorical nature of visual short-term memory

Philip T. Quinlan

University of York, UK

Dale J. Cohen

The University of North Carolina at Wilmington

Author Note

Please address any correspondence regarding the work to Philip Quinlan, Department of Psychology, The University of York, Heslington York, North Yorkshire, YO10 5DD, UK. Email: philip.quinlan@york.ac.uk. We would like to express our sincere thanks to Hao-Ting Wang for help with testing the participants in Experiments 2 – 4 and for generating the additional photo-objects used in Experiments 3 and 4.

Abstract

We conducted a series of recognition experiments that assessed whether visual short-term memory (VSTM) is sensitive to shared category membership of to-be-remembered (tbr) images of common objects. In Experiment 1 some of the tbr items shared the same basic level category (e.g., hand axe): Such items were no better retained than others. In the remaining experiments, displays contained different images of items from the same higher-level category (e.g., food: a bagel, a sandwich, a pizza). Evidence from the later experiments did suggest that participants were sensitive to the categorical relations present in the displays. However, when separate measures of sensitivity and bias were computed, the data revealed no effects on sensitivity, but a greater tendency to respond positively to non-category items relative to items from the depicted category. Across all experiments, there was no evidence that items from a common category were better remembered than unique items. Previous work has shown that principles of perceptual organization do affect the storage and maintenance of tbr items. The present work shows that there are no corresponding conceptual principles of organization in VSTM. It is concluded that the sort of VSTM tapped by single probe recognition methods is pre-categorical in nature.

Keywords: visual short-term memory; principles of organization; conceptual grouping

There is a long standing and continuing debate concerning whether it is sensible to posit separate short-term and long-term memory systems (see e.g., Cowan, 2001; Crowder, 1989, 1993; Endress & Potter, 2014; Melton, 1963). Much of this debate has concerned the retention of verbal materials (see e.g., Roodenrys & Quinlan, 2000), but there is growing interest in whether it is sensible to distinguish visual short-term memory (VSTM) from long-term memory (LTM) (see the recent review by Brady, Konkle & Alvarez, 2011). The present series of experiments assess a primary prediction of the hypothesis that VSTM is separate from and independent of long-term memory, namely, that conceptual knowledge will not influence the ability to retain visual information over short intervals. If conceptual knowledge does influence the ability to retain visual information over short intervals, we can conclude that VSTM is not independent of long-term memory. If, however, such an influence is not found then, the data will provide some support for the functional separation of VSTM from long-term memory.

Whether or not it is sensible to fractionate memory into separate information processing modules has been a somewhat contentious issue (see Cowan, 1995). Here we begin with the distinction between different short-term and long-term stores. This distinction is called into question by two kinds of evidence: (i) data demonstrating that the nature of forgetting is the same over the short- and long-term, and, (ii) data demonstrating that long-term knowledge affects retention of material over the short-term. The first line of evidence relates to the similarity of the forgetting curves observed with remembering over, respectively, the short and long-term. Quinlan and Dyson (2008) presented a review of the data showing that the “recency effect” – the finding that the more recently encountered material is better remembered than material encountered further in the past – is observed for both when short lists of verbal materials are tested after a brief delay (Glanzer & Cunitz, 1966) and when long-term knowledge is tested (see Chap. 10). An example of the latter case is that by Crowder (1993) who tested how well American undergraduates were able to recover a chronological list of the Presidents of the USA. In plotting a serial position curve of the relevant recollections, a prominent recency effect was present: a result that shows that recency effects are not limited to short-term testing conditions. Some have taken the common patterns of recency over brief and longer retention intervals to show that the very same memorial processes operate in all cases (e.g., Nairne, 2002).

The second line of evidence relates to how long-term knowledge can influence the retention of verbal material over the short-term. If there is a short-term store that is insulated from long-term memory, then variables that reflect linguistic knowledge should not influence memory for words over brief intervals. Roodenrys and Quinlan (2000) showed, however, the contrary. They compared serial recall of word lists comprising high and low frequency words, respectively. Although similar serial position curves were obtained for both kinds of lists, performance was overall better with high vs. low frequency words (see also Hulme, Roodenrys, Schweickert, Brown Martin & Stuart, 1997). This evidence conflicts with the hypothesis that there are separate short-term and long-term verbal memory stores that exist and that operate independently of one another.

In addition to the effects of linguistic knowledge on the short-term retention of words, there is also evidence that categorical knowledge can influence short-term memory performance in a similar way. For instance, Poirier and Saint-Aubin (1995) compared serial recall with word lists comprising items all taken from the same semantic category (i.e., sports, musical instruments), with lists comprising items from different semantic categories. Although performance with both kinds of lists produced the same kinds of serial position curves, items from the same semantic category were better recalled than were items taken from different categories. This data suggests that long-term knowledge exerts a significant influence on memory for verbal material over the short term. In the initial report, Poirier and Saint-Aubin (1995) argued that the effects in their data were most indicative of processes operating at recall whereby “taxonomic category could be used as an extra retrieval cue supporting recall” (Poirier, Saint-Aubin, Mair, Tehan & Tolan, 2015, p. 491). In this case, the claim is different from the idea that categorical information operates as an organizational principle in memory because the effect of long-term knowledge takes place only at the point of recall. However, more recently, Poirier et al. (2014) have argued that the current evidence is much more in line with the idea that verbal STM reflects nothing other than temporary activation of the long-term store (cf. Cowan, 2001).

The blurring of the boundary between short-term and long-term *verbal* memory systems suggests that there may be a similar blurring between short-term and long-term *visual* memory systems. However, traditionally, it has been accepted that separate short- and long-term visual stores exist. For instance, the independence of short-term and long-term visual memory systems is central to the modal model of memory as set out by Atkinson and Shiffrin (1968) and it also is foundational to the working memory model of Baddeley and Hitch (1974, and see Baddeley, 2000, for a more recent update). Within this latter theoretical framework, short-term visual memory is discussed in terms of the *visuo-spatial scratch pad* (Baddeley, 2000) – a slave sub-system, separate from others that deal with verbal material, and separate from long-term memory.

The motivation for this kind of separate short-term visual memory system primarily came about from evidence in dual task experiments showing that retention of visual material over the short-term was relatively unaffected by the addition of a secondary verbal task (see e.g., Logie, Zucco & Baddeley, 1990). Moreover, as Hurlstone, Hitch and Baddeley (2014) have pointed out, the distinction between verbal and visual memory systems is underscored by the reports of a double dissociation in neuropsychological patients: Some patients show “impairments in verbal short-term memory, but not visuospatial short-term memory, whereas other patients exhibit the converse pattern of preservation and impairment” (op. cit. p. 341). At the very least, the evidence suggests the existence of a short-term memory system that is dedicated to the retention of visual as opposed to verbal materials.

By analogy with the work on verbal memory, one way to test the independence of short- and long-term visual systems is to examine the influence of long-term knowledge on the retention of visual information over brief intervals. However, there are very few cases where this has been attempted. Konkle and Oliva (2007) reported one notable example. In an initial rating study, photo-images of single real-world objects of various sizes were used and, on each trial, participants were instructed to alter the size of the image so that it was at the “best size to see it”. Generally these ratings were highly consistent over individuals. In this way the, so-called, *normative size* of each the imaged objects was found.

Konkle and Oliva (2007) went on to examine performance in a variant of a same/different paradigm. On each trial an image of an object was presented for 1 s, it was then masked for 200 ms, and then a second image of the same object was presented until response. On Same trials the two images were identical but on Different trials the size of the images differed. Participants were instructed to base their responses on detecting such a size change. Two kinds of Different trials were examined. In the *change towards the normative size* *condition* the size of the second image approached that of the normative size. In the *change away from the normative size* *condition* the size of the second image was displaced away from the normative size. In analyzing *d’* scores, Konkle and Oliva (2007) reported that the size change was much harder to detect in the change to the normative size condition than the change away from the normative size condition. They took this result to suggest the presence of a size distortion effect in memory such that a normalization bias operates in the maintenance of the metrical properties of the images. Such a bias reflects an influence of stored object knowledge on short-term visual memory.

An additional experiment was divided into two parts. In the initial training part, 20 images from the rating study were used. These images were presented to a new sample of participants who were instructed to learn the images prior to being tested in the second part of the experiment. Critically, each image was either too big or too small relative to their normative size. In the final testing part, and on each trial, an image of one of the original objects was presented and the participant had to re-size this to match that shown initially. The findings showed that participants had a clear response bias to re-size the images towards the corresponding normative size. The authors took such errors of re-construction as evidence of a process by which memory for objects is biased towards perceptual norms. Other evidence was cited as being consistent with idea that “memory for the size of a real- world object is influenced by our prior expectations about its size” (Brady et al., 2011; p. 10; e.g., Hemmer & Steyvers, 2009).

We might query whether the very same memory system is being tapped across the same-different and reconstruction experiments given that the time course of study-then-test is very different in the two cases. Moreover, we might also query whether the reconstruction effects are truly a reflection of storage rather than processes concerned with reconstruction itself. That is, the bias towards normative size might come about as a consequence of having to reconstruct the images. Nonetheless, given that the same kind of bias is seen to operate in the two cases this suggests that there is a common cause. It seem reasonable to conclude, therefore, that memory for the *metrical properties* of images of familiar objects can be distorted by stored knowledge and this does provide an example of the influence of long-term memory on the operation of visual short-term memory.

We wish to go further, though, and question whether long-term *conceptual* knowledge influences retention of material over the short-term. An example of how this takes place is typically discussed in the context of Miller’s (1956) classic discussion of *chunking*. Key here is the finding that the recall of collections of spoken letters is improved if they are parsed according to familiar concepts than if they are not so parsed. For example, memory for the letter sequence FBIUSAPHD was improved if pauses were inserted between FBI, USA and PHD (Bower & Springston, 1970). Such a demonstration fits with the idea that conceptual information forms the basis of a principle of organization in STM such that ideas that are conceptually related are grouped together during retention. We may refer to this kind of grouping as *conceptual grouping* as it is conceptual knowledge that governs the formation of groups. Of interest now is whether conceptual content can determine the grouping of items in VSTM. That is, can we find any evidence that items are grouped in VSTM according to a shared identity or common conceptual category?

Although there very little research on conceptual grouping in VSTM, demonstrations of *perceptual grouping* in visual memory have a reasonably long history in the literature (Kahneman & Henik,1977). More recently, Quinlan and Cohen (2012) examined the degree to which short-term memory for simple geometrical colored shapes was affected by common perceptual features. Participants were presented with a brief display containing six colored shapes, and following a brief pause a probe consisting of a single colored shape was then presented. Participants were instructed to respond on a Yes/No basis as to whether the probe had occurred in the memory display. Across different display conditions, single pairs of items, within a given memory display, either shared a common color or a common shape. Moreover on No trials the probe comprised a combination of a color and a shape from different items in the initial display.

The authors identified a *color-sharing* effect, whereby color/shape bindings were better retained if they shared a common color than if they did not. In addition, the data failed to show a corresponding *shape-sharing effect*. Memory for color/shape bindings did not show a similar benefit in cases where the items shared a common shape. There are several possible reasons for the color vs. shape contrast; for instance, it is possible that the colors used were more perceptually distinctive than the shapes. Nonetheless, Quinlan and Cohen (2012) took the color-sharing effect to reflect how color acts as a powerful organizational principle in VSTM They argued that items that share a common color are grouped together and stored as groups in VSTM (see also Morey, Cong, Zheng, Price & Morey, 2015).

The evidence provided by Quinlan and Cohen (2012) reveals the operation of grouping in VSTM based around the shared perceptual properties of the to-be-remembered (tbr) items. Indeed there is a growing body of evidence showing the importance of how color influences memory for tbr items over the short term. Lin and Luck (2009) have documented how tbr items of highly similar colors are better remembered than items of dissimilar colors (see also Sanocki & Sulman, 2011). Peterson and Berryhill (2013) have shown how items that share a common color within a memory display are better remembered than items that possess unique colors – although the benefit was limited to items that both shared a common color and were proximal to one another. Indeed in a number of studies other organizational principles have been examined in the context of VSTM. In some of these color together with other factors have been studied – such as connectedness (van Lamsweerde, Beck, & Johnson, in press) and proximity/good form (Brady & Tenenbaum, 2013). In others the influence of other grouping cues have (see Woodman, Vecera & Luck, 2003, for evidence regarding proximity and connectedness; see Jiang, Chun & Olson, 2000, for evidence regarding contextual orientation cues). There is growing evidence therefore that a fuller understanding of the operation of VSTM can only come about through proper consideration of perceptual grouping and the factors that determine such grouping (see Brady et al. 2011 for further discussion).

Here, we ask whether it is possible to garner evidence showing grouping based on common conceptual category. Such a possibility follows naturally from work reported by Potter and colleagues (Endress & Potter, 2014; Potter, 1976; Potter & Levy, 1969; Potter, Staub, Rado, & O’Connor, 2002). Much of the evidence relates to the recollection of information conveyed in very rapid sequential presentations of lists of photographic images. The evidence from this kind of paradigm suggests that a pictured item’s superordinate category identity can be recovered within 200 ms (i.e., within the time-course of a single fixation, Endress & Potter, 2014). The surface detail of the images are also recovered within this time course but, whereas conceptual content can be retained in a relatively durable format, the surface details are retained in a relatively fragile form and are prone to rapid forgetting. This supports the separation of short-term stores for perceptual and conceptual information, respectively. Of key interest here is whether the rapid extraction of such categorical information supports conceptual grouping in VSTM.

One previous study provides some data that are of direct relevance to current concerns. In their change detection experiment, Wong, Peterson and Thompson (2008) examined short-term memory for arrays of visual objects. All displays contained either two or four images and the initial memory display was presented for either 1 s, when the display contained two images, or 2 s when the display contained four images. A random line mask was then presented for 1 s and a probe display was then presented. When there was a change in the probe display, one of the previous images was substituted with a new image. Participants made change vs. no change responses following the onset of the probe display.

In addition to the manipulation of memory set size, they also manipulated the number of kinds of stimuli contained in the display. Critically, the set size four displays could comprise four objects from the same category or two objects from two categories. In their first experiment they used a combination of faces and houses and a central result was that memory was better when two objects from two different categories were presented than when four objects from the same category was presented. By examining performance with a range of different categories, however, they discovered that this multiple category benefit only occurred when one of the categories was that of faces. Wong et al. (2008) examined combinations of faces, houses, bodies, watches and butterflies. Overall therefore these data were taken to argue for the special treatment for faces in memory, but there was no evidence of any further effects of category membership in VSTM.

Although Wong et al. (2008) referred to their images as coming from different kinds of object categories, from the examples given, the visual nature of the images varied dramatically across the categories. That is, photographs of houses in real world scenes, photo-images of individual butterflies and watches, line drawings of human figures, and, images of morphed faces, were used. Evidently, therefore there are both categorical and clear perceptual differences across the image sets and as a consequence it is difficult to know whether the effects reflect conceptual or perceptual factors. Nonetheless, the data are of interest because they do not reveal any general influence of object category when the tbr images were presented simultaneously. In this regard, there are some grounds for caution in predicting a category effect in the present experiments despite the other evidence for the rapid retrieval of category identity when the images a present rapidly and sequentially (see e.g., Potter, 1976).

Here, we report a series of experiments that assess whether conceptual content plays a critical role in VSTM. If there is a short-term visual store that is insulated from long-term memory, then variables that reflect conceptual knowledge should not influence memory for images over brief intervals. A default hypothesis is that items that share a common category ought to be better retained than items that do not. This follows from a consideration of how perceptual grouping/chunking operates in VSTM. A basic finding is that items that are grouped together are better remembered than are items that do not so group. We wished to explore whether such grouping effects can be shown to occur on the basis of shared category membership. This in turn depends on the idea that the category identity of the tbr items is recovered rapidly (cf. Potter, 1976) and that this operates as an organizing principle in VSTM. Any such effects of conceptual grouping follow naturally from the idea that VSTM is nothing other that activation of items in long-term memory.

In Experiment 1 we compared cases where items share a common identity - for instance two images of different, hand axes presented amongst other unrelated images – with cases where their identity is unique in the display. If the short-term visual store is not insulated from long-term memory, we predicted that memory for items that share a common identity (henceforth, *category items*) would be better than memory of *unique items*.

For ease of exposition, we will draw a distinction between shared identity and shared membership of a common higher-level category. This maps onto a terminological distinction between *basic level* and *higher-level categories* (after Rosch, 1975). Two different hand axes belong to the same basic level ‘hand axe’ category, but they also belong to the higher-level ‘hand-tool’ category. Experiment 1 addresses the possible influence of basic level categories on the organization and storage of pictures in VSTM. The remaining experiments address issues concerning higher-level categories. In these latter cases, instances from different basic level categories are used that belong to the same higher-level category - for instance a hand axe and a hammer.

General Methods

We used a single probe recognition task in all experiments. On every trial, an initial memory display containing six tbr images was presented briefly. Following a brief pause, a single probe image was presented. Participants were then to decide whether the probe had been present in the display.

All participants were tested individually in one of two laboratories. The experiments (except Experiment 1) were controlled by Eprime 2 scripts (Schneider, Eschman, & Zuccolotto, 2002) running on stand-alone PCs. Lab 1 housed a sound attenuated chamber and Lab 2 contained a partitioned testing cubicle. In both cases, a computer monitor (Viglen Model 1769ME) was situated on a table (centered at eye-level) approximately 57 cm from the edge of the table. Experiment 1 was controlled by bespoke Python script. Finger response keys (in Lab 1) and a keyboard (in Lab 2) were used to collect responses. The controlling computer was located outside of the testing spaces in both laboratories. Participants in Experiments 1 were all tested in Lab 2 and the remaining participants were tested in Lab 1.

Experiment 1

*Method*

*Participants*

Participants were students currently enrolled at The University of York. Participants were recruited via a university-wide web portal and either received £4 or course credit for taking part. The eventual sample size was 25 comprising 20 females and 5 males. All 25 were right-handed (mean age = 21.1 yrs, stdev = 5.3 yrs). In arriving at the final sample, 2 participants were replaced due to poor levels of accuracy at the task (i.e., they attained less than 53% correct overall). The mean overall level of accuracy across the final sample was 67.2% correct (min = 61.7% correct). All of the participants had normal or correct-to-normal vision and gave informed consent prior to the experiment.

*Stimuli, Design and Procedure*

The stimuli throughout were photo-images of naturally occurring objects. The images were taken from the database of photo-objects used by Konkle, Brady, Alvarez and Oliva (2010). This database comprises 200 sub-directories each containing 17 images from a given basic level category (e.g., doll, bike, frisbee, etc.). Eighteen of the categories were deemed unsuitable for a variety of reasons (e.g., American road signs were deemed culturally inappropriate) and were not used.

Six conditions were generated and defined relative to the composition of the memory display. Every memory display contained six tbr images. In Condition (Aa)bcde two item (i.e., ‘A’ and ’a’) were taken from the same category and ‘b’, ‘c’, ‘d’, ‘e’, were items selected from four other different categories. ‘A’ signifies that the probe was taken from the same ‘A’ category as the corresponding memory item and items within the brackets instances of the same category type. An implication behind this condition is that the probe is testing the memory of the category items.

Using this kind of notation there were four other conditions, namely, Condition (aa)Bcde in which there was a pair of category items in the memory display but the probe tested memory for a unique item, Condition (Aa)(bb)cd in which there were two pairs of category items and the probe tested the memory for one of these, Condition (aa)(bb)Cd in which there were two pairs of category items and the probe tested the memory for one of the unique items, Condition (Aa)(bb)(cc) in which there were three pairs of category items, and finally Condition Abcdef in which there were six unique items.

In designing the displays in this way, each condition had its own cases of Yes and No probes. Importantly, in all cases the probe was taken from the same category as a memory item. Thus, a strategy based on simply naming the tbr items would not work because the probes always shared a category label with a corresponding tbr item.

 The display area of the images on the screen was divided into a virtual 6 x 6 array (approximately 24o x 20o in extent). The placement of the memory images (each approximately 3o x 3o in extent) was determined at random within this area. Prior to each trial, each display position was sampled without replacement and the location of the probe was chosen not to correspond with the location of any of the memory images.

At the start of each trial, the images were selected at random from the database given the display constraints of the current condition. Each trial started with the presentation of a central fixation ‘+’ for 1000 ms, followed immediately by a memory display for 750 ms. The memory display was followed by a blank interval of 500 ms, and then the probe was presented until the computer detected a keyboard response. Participants were instructed to press a ‘1’ on the keyboard to signify that they thought the probe had been present in the display and ‘2’ if they thought it had not. The word ‘Correct’ or ‘Error’ was then presented for 500 ms prior to the start of the next trial as contingent feedback. Figure 1 provides a schematic representation of the events on a trial in the experiment.

Each participant was tested individually. Prior to the start of the experimental trials 12 practice trials were presented. There then followed 3 blocks of 120 trials. Positive and negative cases for each of the six conditions were equally represented in each block. The order of trials, within a block, was randomly chosen on a participant-by-participant basis.

*Results and Discussion*

The average proportion level of accuracy across the participants was .67. The individual accuracy scores for each participant were converted into *d’*s and these scores were entered into a one-way repeated measures ANOVA in which condition (the six conditions) was entered as a fixed factor and participants was entered as a random factor. The analysis revealed that the main effect of condition failed to reach statistical significance, *F*(5, 120) = 0.882, *p* > .05, *𝞰2* = .035. Summary data are presented in Figure 2 and, as can be seen, sensitivity did not vary across the conditions of interest.

Overall therefore there was no evidence that participants were differentially sensitive to the categorical relations in the memory displays. That is, memory for items that shared a basic level category was no different from memory for unique items. In this respect the data failed to show any evidence of grouping based on shared identity – that is there is no evidence of conceptual grouping in the present data. In this regard, these data accord with the similar null effects of grouping reported by Wong et al. (2008). The remaining experiments follow up on this finding, so as to ascertain, what if, any effects of category membership can be located within VSTM.

Experiment 2

 Experiment 1 suggests that categorical knowledge does not influence VSTM. Comparing across that experiment and the previous work by Quinlan and Cohen (2012), it seems that, whereas the organization of information in VSTM is affected when items share a common color, similar grouping does not occur for shared identity. In this regard, it seems that the memory system being tapped in the task is pre-categorical in nature (Broadbent, 1958). We followed-up this experiment with additional studies in which different timings and display configurations were used. Summaries of the results from two of these studies are included in Appendix A, for illustrative purposes. The results were robust and consistent: there was no evidence of conceptual grouping in any of the additional studies.

From a consideration of this body of evidence, we therefore conclude that the null effect in Experiment 1 reflects the more general fact that VSTM is pre-categorical in nature. Although the identity of the tbr images may well be rapidly recovered (cf. Potter at al., 2002), it seems that this information plays no further role in the storage and maintenance of these items in VSTM. On these grounds, we adopted a new approach to the basic issue and developed a different variation of the single probe recognition task. Our aim was to examine whether *any* evidence could be found for conceptual grouping on the basis of instances sharing a common higher-level category (e.g., different kinds of Food – a bagel, a sandwich, etc.).

Initially Experiment 2 was carried out in a bid to examine the degree to which higher-level category information is recoverable from the kinds of brief displays used in VSTM single probe recognition studies. In this experiment, each display contained six images. Each of the images was of an instance from a different basic level category, but all instances belonged to the same higher-level category. In this case the primary index of performance was accuracy on the No probe trials. Three kinds of No probe trials were used: a *New* probe, an *Other* probe, and an *Alternate* probe. The New probe was an item completely unrelated to the memory items. The Other probe was taken from the same higher-level category as the memory items but its corresponding basic level category was not represented via one of the memory items. The Alternate probe was taken from a basic level category shared with one of the memory items.

Nine higher-level categories were chosen and instances from seven constituent basic level categories were selected for each of these. When a higher-level category was chosen, six basic level categories were then selected at random and instances from these were used as the tbr images for that trial. On the Other probe trials the probe image was taken from the final (remaining) basic level category. As the nine higher-level categories were sampled repeatedly over the trials it is assumed that participants became familiar with these and the corresponding basic level constituent categories. As each basic level comprised 17 instances and because these were sampled with replacement over trials, it is also possible that a given image was also presented more than once (though the probability of this happening was small).

*Method*

*Participants*

The same participants were tested in what we term Experiments 2 and 3. The order of testing of these two experiments was counterbalanced across participants. Twenty-four University of York students (18 women and 6 men, *M* = 23.0 years, *SD* = 2.27, age range: 18– 31 years) participated in the experiment. All of the participants had normal or correct-to-normal vision and gave informed consent prior to the experiment. Participants either received payment or course credit for completing the testing.

*Stimuli, Design and Procedure*

In choosing the stimuli, the aim was to select images of basic level instances of the same higher-level category. A design constraint was that each higher-level category needed to contain seven different basic level categories. For instance, the higher-level category Food comprised the basic level categories bagel, loaf, cake, cookie, muffin, pizza, sandwich (see Appendix B for a complete listing of the nine key categories). Each basic level category contained images of 17 different instances.

Each memory display contained six images and each image was selected at random from the current higher-level category – for instance a bagel, a loaf, a cake, a muffin and a pizza. The experiment contained four conditions, each defined relative to a different kind of probe item. In the Same condition the probe image was one of the actual tbr items. In the Alternate condition a different instance from one of the six basic level categories was presented – so that if cake1 was presented as a tbr image then cake2 acted as the probe. In the Other condition the probe image was taken from the seventh basic level category that had not been included in the memory display. For instance, the probe was an item of food not previously represented in the memory display (i.e., a sandwich). Finally in the New Condition the probe was an item unrelated to any of the category images shown in the memory display. Across the conditions, the Same condition trials demanded a Yes response and the remaining three conditions demanded a No response.

Prior to the start of the trial images were selected at random given the constraints of the particular condition. In addition, further constraints were applied such that clear associates with the selected category were excluded from the selection. Lists of obvious associates for each of the nine key categories were compiled on the basis of the intuitions of the first author.

Memory displays comprised 6 images arrayed equidistant from one another around a virtual circle (~9o in diameter). The individual images were approximate 2.5o x 2.5o in extent. The probe image was presented centrally. An experimental trial began with a fixation cross for 1000 msec. The memory display was then presented for 750 ms, followed by a blank screen of 500 msec. The probe image was then presented until the computer detected a key press response. Figure 3 provides a schematic representation of the events on a trial in the experiment.

Participants were tested individually. Twenty-seven practice trials were followed by four experimental blocks each comprising 54 trials. Overall across the experimental trials, there were 108 that demanded an Old response (in same condition) and 36 New trials for each of the other three conditions.

*Results and Discussion*

In all other experiments measures of *d’* were analyzed, but in this case, there was only one Yes condition, hence the hit rate is the same for all No conditions. Therefore simple accuracy scores were analyzed. The average overall proportion correct over the participants was .77. Primary interest was with how performance varied over the three kinds of No probe trials. Accuracy scores in these conditions were converted to proportion correct and then arscin transformed according to the advice of Winer (1971). These transformed scores were then entered into a one-way, repeated measures ANOVA in which condition (alternate, other and new) acted as a fixed factor and participants acted as a random factor. This analysis revealed that the main effect of condition was statistically reliable, *F*(2, 46) = 72.56, *p* < .001, *𝞰2* = .759. Further Bonferroni corrected, pair-wise comparisons revealed that participants were most accurate in the New condition, *t*(23) = -11,24, *p* < .001, for the comparison with the Alternate condition, *t*(23) = -6,25, *p* < .001, for the comparison with the Other condition, and least accurate in the Alternate condition, *t*(23) = -6,29, *p* < .001, for the comparison with the Other condition. Figure 4 provides a graphical illustration of the proportion correct scores in all four conditions.

The results suggest that the participants recovered category information from the memory displays, as is witnessed by the performance differences across the various kinds of No probe trials. The tendency to make a false alarm response varied according to the conceptual similarity of the probe to the memory items. This tendency was greatest when the probe was a different instance of a basic level category represented in the display (i.e., the Alternate condition, e.g., cake2 probe when cake1 was the corresponding memory item). It was reduced when the probe was an instance of the same higher-level category as the tbr items but did not share the same basic level category as any tbr item (i.e., the Other condition). Finally, it was least when the probe was an instance of a different higher-level category than the tbr items (i.e., the New condition).

The fact that participants also tended to err in the Other condition is of some interest. This might be taken to show that participants recovered information about the higher-level category represented in the memory – such as Food – that is, they extracted this kind of gist from the briefly presented memory display. As a consequence, when any item of Food acted as a probe they tended to respond Yes.

Alternatively, participants may have been simply responding to the co-occurrence of basic level categories across the displays. Across the displays, cakes and bagels tended to co-occur. Therefore, if cake1 is retrieved from the memory display, the participant may be biased toward responding positively to a bagel1 probe even though bagel1 was not stored. Evidence for this kind of tacit learning of the statistical regularities in similar VSTM paradigms has been well documented by Brady, Konkle and Alvarez (2009).

We conducted four further experiments to determine whether the results of Experiment 2 are indicative of long-term memory influence or simply response bias. In order to do this, tasks were designed in which separate measures of sensitivity and bias could be computed for the key display conditions.

Experiment 3

 The data from Experiment 2 provide some grounds for thinking that higher-level category information might operate as an effective grouping principle in VSTM. To test this idea Experiment 3 was developed on the basis of aspects of both Experiment 1 and Experiment 2. Now, in every display, three images from a common higher-level category were presented and each was taken from a different constituent basic level category (e.g., bagel, loaf, sandwich, namely, the *category items*). The remaining three images were chosen at random from unrelated other basic level cases (i.e., the *unique items*). Similar probe trials to the previous Alternate and Other trials in Experiment 2 were used and both category and unique items were probed in a similar fashion to that described in Experiment 1.

*Method*

The general method was the same as reported for Experiment 2. The only critical difference was with respect to the display composition and categories used.

*Rating study*

 One concern with the previous study was that the choice of the materials was governed by the intuitions of the first author. To standardize the materials in the present experiment, a random sample of 26 students were presented with 13 category labels and asked to write down 10 instances of these categories. These responses were then collated and the four most common instances in each category were noted. One category (Sports equipment) was excluded because three of the four most frequently rated items were reported by less than a third of the participants. One object with low perceptual variety (i.e. rice) and one with high perceptual similarity with another selected item (i.e. pen and pencil) were excluded and the next most frequently reported item in the corresponding category acted as a replacement. The detail of 12 categories, items and their reported frequencies are in Table 1.

A consequence of this rating exercise was that some of the commonly reported instances were not included in the extant database. These instances are picked out in Table 1 and new photos had to be sourced for these. In total 17 new items for each of 26 basic level categories were sourced and added to the database.

*Memory Task*

*Participants*

 The participants were the same as those in Experiment 2.

*Stimuli, Design and Procedure*

 The critical differences across Experiments 2 and 3 relate to the composition of the memory displays and the types of probes used. All memory displays contained six images, but three of the images (the category items) were taken from the same higher-level category. Each tbr image was selected at random (given the constraints of the particular trial display condition). Images were allocated to display locations at random. From Table 1 the higher-level category might be Food and the three images would be chosen at random from the four possible instances listed under Food (e.g., chocolate, pasta and pizza). The remaining three items (these constituted the unique items) were chosen at random from the other images in the database. A schematic representation of the sequence of event on a trial is included in Figure 5.

In the Same-In condition, the probe was one of the original category items. In the Same-Out condition, the probe was one of the original unique items. In the Alternate-In condition, the probe was a new item taken from one of the sub-categories used in the memory display (in the current example, a new image of chocolate). In this way, the probe was a new item not included in the memory display. In the Alternate-Out condition, the probe was a new item taken from one of the unique item’s sub-categories used in the memory display (in the current example, a new image of one of the non-food items included in the display). In this way, the probe was a new item not included in the memory display. In the Other condition, the probe was taken from the higher-level category instance that had not been selected as a tbr item (in the current example ‘breadloaf’). Finally in the New condition, the probe was an unrelated item taken from the remaining database photos.

Participants were tested in individually in a single testing session. Initially there was a block of 12 practice trials and this was followed by three blocks of 96 experimental trials. Trials types appeared in random order within a block. This resulted in 72 trials in same–in and same–out condition, and 36 trials for each of the other four conditions.

*Results and Discussion*

 The average proportion correct over the participants was .73. Proportion correct scores were computed for each condition (on a participant-by-participant basis) and summaries of these data are shown in Figure 6. Similar to Experiment 2, accuracy increased monotonically across the Alternate, Other and New conditions. To determine whether the effect of display condition described above was the result of the influence of long-term memory or simply a response bias, sensitivity scores (*d’s*) were computed for the Same-In and Same-Out conditions. When calculating *d’s*, for the Same-In and Same-Out conditions, the false alarm rates were computed from the Alternate-In and Alternate-Out conditions, respectively. The average *d’s* were 0.877 for the Same-In and 0.838, for the Same-Out condition (see Figure 7a). A paired sample *t* test showed that sensitivity was equivalent across these two conditions, *t*(23) = .486, *p* > .05, two-tailed test. Using the methods discussed by Rouder, Speckman, Sun and Morey (2009), the associated Scaled JZS Bayes factor was found to be 4.183 in favor of the null (scale r = 0.707).

Equal sensitivity across the Same-In and Same-Out conditions, suggests that long-term memory did not influence recognition. Rather, the difference in percent correct scores most likely arose because of differences in response bias. Therefore, we computed measures of response bias for the two conditions in terms of *c* (Macmillan & Creelman, 1991). The average *c* value for the Same-In condition was .158 and for the Same-Out condition it was .405 (see Figure 7b). A paired sample *t* test showed a statistically reliable difference across the two conditions in terms of response bias, *t*(23) = -5.93, *p* < .001, two tailed test. Participants expressed a much stronger bias to respond Old on the Same-Out trials than on the Same-In trials.

 The data clearly reveal that although participants do recover some information about common categories that are represented in memory, this does not facilitate the storage/retention of the tbr items. This evidence provides a clear example of “post-perceptual guessing” (Brady et al., 2009). That is, the likelihood of responding New to a particular probe was determined by whether it was from the same higher-level category as that of the category items. Participants were far more likely to respond New to a probe associated with the category items than with a unique item in the memory display.

Experiment 4

 In Experiments 1-3, we found no evidence for an influence of conceptual knowledge on the ability to recognize items from a briefly presented visual display. Nonetheless, the response bias revealed in Experiment 3 demonstrates that participants are extracting and using category information from the displays. Experiment 4 was our attempt to push this finding as far as possible. Here, instead of relying on a participant’s tacit knowledge of the display composition, we provided, prior to each trial in the cuing condition, an explicit pre-cue as to what kind of items to expect in the memory display. The pre-cue was the visual presentation of the higher-level category name prior to the trial onset. The thinking here was that the category name would prime the long-term memory representation of the category constituents and that this would in turn operate as an effective means of organizing the up-coming tbr items. Performance in the cuing condition was tested against a no-cue condition (as used in Experiment 3) in which no category pre-cues were presented.

*Methods*

*Participants*

Twenty-four University of York students (15 women and 9 men, *M* = 20.2 years, *SD* = 2.1, age range: 18– 25 years) participated in the experiment. All of the participants had normal or correct-to-normal vision and gave informed consent prior to the experiment. Participants either received payment or course credit for completing the testing.

*Design and Procedure*

A key difference between Experiments 3 and Experiment 4 was the presentation of a category label at the start of half of the trials. The category label named the higher-level category of the three category items in the memory display and this name was displayed centrally for 1s at the start of each trial. On the No Prime trials the label was replaced with a blank screen. In addition, only four conditions from Experiment 3 were re-tested: the Same-In, Same-Out, Alternate-In, and Alternate-Out conditions. There were equal number of prime and no prime trials tested for each of these four conditions.

In each testing session a block of 24 practice trials preceded three blocks of 96 experimental trials. This resulted in 36 observations in each of four Prime and No Prime conditions.

*Results and Discussion*

 The average overall proportion correct over the participants was .67. Proportion correct scores were computed for each condition and on a participant-by-participant basis and to separate out sensitivity from response bias, corresponding accuracy scores were transformed in measures of *d’* and *c*. Graphical summaries of these data are provided in Figure 8. The *d’* scores were analyzed using a 2 x 2 repeated measures ANOVA in which prime condition (Prime vs. No Prime) and display condition (Same-In vs. Same-Out) were entered as fixed factors and participants was entered as a random factor. This analysis revealed that none of the tests resulted in a statistically reliable effect, all *Fs* < 1. The Scaled JZS Bayes factors were computed for the Prime and No Prime conditions respectively (after Rouder et al., 2009). The corresponding values were 2.969 for the Prime condition and 4.489 for the No Prime condition.

 In contrast, the corresponding analysis of the *c* scores revealed that the main effect of display condition was statistical reliable, *F*(1, 23) = 16.51, *p* < .001, *𝞰2* = .42, *F* < 1, for both the main effect of prime condition and the prime condition x display condition interaction. The reliable main effect of display condition reveals the same pattern of response bias as that found in Experiment 2: Participants expressed a stronger bias to respond Old on the Same-Out trials then the Same-In trials.

 The patterns of performance across the Prime and No Prime conditions were essentially the same: Memory was no better for grouped than ungrouped items. Priming had no material effect on participants’ memory for the grouped items. Nonetheless, participants exhibited the same response bias demonstrated in Experiment 3, regardless of priming. That is, they were more inclined to respond Old to a probe related to the unique items than the category items.

Experiment 5

 Experiments 1-4 present fairly strong evidence that conceptual knowledge has no influence on the ability to recognize items in a briefly presented visual display. This final experiment addresses any remaining concerns about possible perceptual differences that may cut across the sorts of categories that we have examined. It is possible that the experiments have been compromised by subtle yet systematic visual differences across the various kinds of category instances that have been used. For instance, item similarity has, perhaps, deleteriously affected memory for the category instances on the grounds that within-category items share more common features than between-category items. Therefore memory for particular category instances has been compromised by the presence of other highly similar tbr instances. Under this hypothesis, conceptual knowledge aids recall, but the visual similarity of within-category items exactly cancels out that benefit. Although such an exact cancelation effect is unlikely, here we assess the *within-category detriment* hypothesis by systematically varying the number of category instances within a given display. The within-category detriment hypothesis predicts that memory for a given category instance ought to scale inversely as a function of the number of other category instances in the display.

 As a consequence in the final experiment we systematically varied, across trials, the number of grouped items present in the memory display. Four main conditions were designed. In the *category condition* all six images were taken from the same category and in the *random condition,* instances from six different categories were chosen. For both of these this conditions Same-In and Alternate-In cases were generated. In addition a *Group 2 condition* and a *Group 4 condition* were designed. In the Group 2 condition two of the images were from the same category and in the Group 4 condition four of the images were from the same category. The remaining images were chosen at random from other categories. For the Group 2 and 4 conditions Same-In, Same-Out, Alternate-In, and Alternate-Out cases were designed. The within-category detriment hypothesis predicts the lowest recall sensitivity in the category condition, followed by the Group 4 and Group 2 condition, with the highest recall sensitivity in the random condition.

*Methods*

*Participants*

Twenty-four University of York students (15 women and 9 men, *M* = 20.8 years, *SD* = 2.0, age range: 18– 25 years) participated in the experiment. All of the participants had normal or correct-to-normal vision and gave informed consent prior to the experiment. Participants either received payment or course credit for completing the testing.

*Design and Procedure*

 Twelve categories of items were sourced such that each category contained 6 instances. Sports Equipment was substituted for Shoes in this final experiment so as to avoid any issues regarding Shoes vs. Clothes.

The presentation conditions were the same as those used in the no-prime condition in Experiment 4. In each testing session a block of 12 practice trials preceded four blocks of 120 experimental trials. This resulted in 40 observations in each of the conditions of interest.

*Results and Discussion*

 The average overall proportion correct over the participants was .64. Proportion correct scores were computed for each condition and on a participant-by-participant basis. As before separate analyses are reported for the *d’* and *c* scores.

*d’s*

Initially the *d’* scores were examined to see whether sensitivity scaled (inversely) as function of number of grouped items in the displays. To test for such a relation the data from the random, Group 2 Same-In, Group 4 Same-In and category conditions was considered. Respectively, across these conditions, there were 0, 2, 4, and 6 category items present in the displays. Table 2 provides summary statistics for the conditions of interest. The corresponding *d’*s were entered into a one-way repeated measures ANOVA in which condition (random, Group 2 Same-In, Group 4 Same-In and category) was entered as a fixed factor and participants was entered as a random factor. The analysis revealed that the linear contrast failed to reach statistical significance, *F*(1, 23) = 1.08, *p* > .05, *𝞰2* = .045. There was no evidence therefore that sensitivity scaled inversely with the number of grouped tbr items. Moreover, sensitivity did not change across these conditions, *F* < 1.0, for the main effect of condition,

To maintain parity with the previous methods, the scores for the two grouped conditions were entered into a 2 x 2 repeated measures ANOVA in which grouping condition (Group 2 vs. Group 4) and display condition (Same-In vs. Same-Out) were entered as fixed factors and participants was entered as a random factor. Figure 9 provides a graphical summary of these data (a) together with those for the *c* scores (b). In this analysis none of the tests reached statistical significance, *F* <1.0, for the main effects of grouping condition and display condition, *F*(1, 23) = 1.41, *p* > .05, *𝞰2* = .06.

*c scores*

 The same kinds of analyses were carried out on the *c* scores as those for the *d’s*. Comparing across conditions, the ANOVA revealed that the linear contrast did reach statistical significance, *F*(1, 23) = 21.97, *p* < .001, *𝞰2* = .49, in the analysis (see Table 2 for summary c scores). This result does show that the bias to say Old decreased as the number of category instances in the display increased.

 Analysis of the *c* scores for the grouped conditions revealed that both the main effect of grouping condition, *F*(1, 23) = 14.92, *p* = .001, *𝞰2* = .39, and display condition, *F*(1, 23) = 27.96, *p* < .002, *𝞰2* = .55, reached statistical significance, F < 1.0, for the grouping condition x display condition interaction. This pattern is line with the previous bias effects.

 At a general level, the results are line with those previously presented. Measures of memory (*d’* scores) have failed to show any systematic differences according to the category composition of the displays. More particularly, sensitivities did not vary inversely with increases in the number of grouped items in the displays. However response biases did vary across the display conditions. The participants’ tendency to respond Old decreased as the number of grouped items in the displays increased. This kind of bias again shows that participants were able to recover some categorical information from the tbr items but that this information did not benefit item memory.

Experiment 6

 A final experiment addresses a concern raised by the fact that all of the experiments reported here have used a relatively short retention period of 500 ms. A question therefore remains about what happens when a longer delay is introduced between the offset of the memory display and the onset of the probe. Perhaps any possible effects of shared category only emerge after a longer delay? To examine such a possibility the final experiment used a retention interval of 1 sec (a duration used previously in studies of perceptual grouping in change detection tasks, see e.g., Jiang et al., 2004).

*Method*

 Experiment 6 is a replication of Experiment 3 with a new sample of participants

*Participants*

 Twenty-four University of York students (20 women and 4 men, *M* = 21.1 years, *SD* = 2.7, age range: 18– 30 years) participated in the experiment. All of the participants had normal or correct-to-normal vision and gave informed consent prior to the experiment. Participants either received payment or course credit for completing the testing.

*Results and Discussion*

 The average proportion correct over the participants was .69. Proportion correct scores were computed for each condition (on a participant-by-participant basis) and summaries of these data are shown in Figure 10. The average *d’s* were 0.671 for the Same-In and 0.749, for the Same-Out condition (see Figure 11a). A paired sample *t* test showed that sensitivity was equivalent across these two conditions, *t*(23) = -0.771, *p* > .05, two-tailed test. The corresponding Scaled JZS Bayes factor (Rouder et al., 2009) was 3.560. The average *c* value for the Same-In condition was -.009 and for the Same-Out condition it was .245 (see Figure 11b). A paired sample *t* test showed a statistically reliable difference across the two conditions in terms of response bias, *t*(23) = -5.27, *p* < .001, two tailed test. Participants expressed a much stronger bias to respond Old on the Same-Out trials than in the Same-In condition. As in all previous cases there was no statistically reliable difference in sensitivity scores for the category and unique items. However there was a reliable difference in response bias.

 To examine more closely any differences in performance as a consequence of the memory retention period, cross experiment comparisons were carried out, separately, on the *d’* and *c* scores for Experiments 3 and 6. In both cases a 2 x 2 split plot ANOVA was used in which experiment (between-participants) and display condition (within-participants) were entered as fixed factors and participants acted as a random factor. The analysis of the *d’* scores failed show a statistically reliable difference in any test, *F*s < 1.0, for the main effect of display condition and the display condition x experiment interaction; *F*(1, 46) = 2.89, *p* > .05, *𝞰2* = .06, for the experiment main effect. In contrast, only the main effect of display condition reached statistical significance, *F*(1, 46) = 61.81, *p <* .001, *𝞰2* = .57, in the analysis of the *c* scores, *F* < 1, for the display condition x experiment interaction. There was also an indication that the response bias was overall smaller in the Experiment 6 than Experiment 3 although the main effect of experiment just failed to reach statistical reliability, *F*(1, 46) = 3.84, *p* = .056, *𝞰2* = .08.

In sum, the same difference in response bias was found in Experiments 3 and 6 in which the retention interval was lengthened across these two experiments, there was no associated change in sensitivity across the experiments. There was however some indication that the overall response bias reduced as the retention interval increased.

General Discussion

 Our data support the view that although some identity information and higher-level category information is recovered from brief displays containing photographic images of real world items, it seems that this plays no role in the organization of tbr items in VSTM. We have shown that although memory for particular instances is not influenced by the presence of other items from the same category, participants are sensitive to such categorical information when they are probed about their memories. We have shown repeatedly statistically reliable effects of response bias but no effects in the *d’* scores. Participants have repeatedly shown a greater bias to respond Old if probed with a unique item than an item that shares the category with the category items. Moreover, this bias scaled according to the number of category items in the displays.

 In Experiment 1 there was no evidence that memory of any particular instance was influenced by the presence of shared tokens of a common type: for example, memory for the image of a hand axe was unaffected by the presence of another hand axe in the display. We take this as an indication that item identity did not operate as an effective organizational cue in VSTM. However, it might also be argued that the failure to find such an effect shows that the system is, essentially, insensitive to the presence of common shapes. Arguably, items of a common identity are highly visually similar. Therefore if common shape does operate as a perceptual organizing principle in VSTM then it would be expected that items of the same type would group together. However there was no evidence for this and taken together with the null effect of common shape reported by Quinlan and Cohen (2012) this is suggestive of a functional constraint on VSTM.

Indeed future research might address the issue of why color is so potent a grouping principle in VSTM and yet shape appears not to be. This is not to argue that color provides the sole basis for grouping in VSTM. As noted earlier, the work of Woodman et al. (2003) and Jiang et al. (2000) provide clear evidence for grouping by other ‘spatial and ‘structural’ factors. Nonetheless, there is now very convincing evidence that color provides a very powerful cue to grouping in VSTM (Brady & Tenebaum, 2013; van Lamsweerde, et al. in press).

 Experiment 2 provided some evidence for a possible role of higher-level category information insofar as participants’ false alarm rates were graded according to the probes relation to the tbr items category. However, according to an alternative view, these data might actually reflect nothing other than the fact that participants were simply sensitive to the co-occurrence of particular kinds of images in the displays. These two alternatives were next tested. In Experiment 3 displays contained both grouped and ungrouped items whereby grouped items shared a common higher-level category. The data very clearly showed no difference in sensitivity to category vs. unique items but there was a clear response bias to responding Old to unique item probes rather than items related to the category.

 In Experiment 4 in the prime conditions a category pre-cue label appeared at the start of every trial alerting the participant to the actual higher-level category defining the grouped items. The data failed to show any sensitivity effects of category primes but the same kind of response bias was revealed for both prime and no prime conditions. Participants were simply more inclined to respond Old to a probe associated with the unique than the category items. This is the very same response bias to that shown in Experiment 3.

 In Experiment 5 the number of grouped items contained in a display was varied across conditions. An aim was to rule out an alternative account of the findings that item similarity may have been responsible for the failure to finding any effects of categorical grouping on memory. If the category items are visually confusable then this would explain the failure to find any such effects. However, yet again there was no reliable difference in sensitivity scores across the conditions and more particularly there was no evidence that memory for the category items scaled inversely as their number within the displays increased. Nonetheless, the same response bias was found as before and there was evidence that this did scale according to the display composition. Participants’ performance did vary according to display composition but only in so far as they were more prepared to make an Old response as the number of unique items increased across the displays. Yet again therefore there is clear evidence that participants were sensitive to the category composition of the displays but this had no consequence on their ability to retain items in memory.

 The final experiment tested a residual concern about whether the effects would generalize to testing conditions in which a longer retention interval was used. As a consequence, the retention interval was increase from 500 ms to 1000 ms and performance was re-tested. The results were generally clear cut and consistent with the previous findings. Participants showed no difference in their sensitivity to the category and unique items, but the same response bias as shown previously was replicated.

*General Theoretical Implications*

On the basis of this overall pattern of findings we conclude that VSTM is pre-categorical in nature. That is, whereas entry into the store and maintenance in the store are demonstrably governed by the perceptual characteristics of the tbr items (see e.g., Quinlan & Cohen, 2012), such operations are *not* sensitive to the conceptual content of the tbr items. Such a conclusion fits relatively comfortably with the idea that there is a separate short-term visual memory system that is functionally independent of long-term memory (cf. Lin & Luck, 2012). In broad terms, the data fit both with the modal model and more particularly with the working memory conception of memory (Baddeley, 2000). The present data also provide evidence that converges with that reported by Wong et al . (2008). They too concluded that the storage and maintenance of items within VSTM is unaffected by the categorical nature of the tbr items.

 Even though there is no evidence here that conceptual content acts as an organizing principle in VSTM, this is not to deny that an item’s identity is rapidly recovered or that there are some kinds of short-term memory systems that are sensitive to conceptual content. The work of Potter and colleagues (Endress & Potter, 2014; Potter, 1976; Potter & Levy, 1969; Potter et al., 2002) provides intriguing evidence regarding these kinds of representations and processes. What the present data suggest is that categorical information does not operate as an *organizing principle* in the maintenance of information in VSTM. It might be argued that the previous findings regarding the rapid recovery of item identity reflects the operation of mechanisms that are dedicated to the processing of single objects one at a time and that different operations are invoked when multiple objects are presented. Such a speculation however remains to be explored in future work.

The current findings are also if interest when contrasted with those relevant to the retention verbal materials over the short-term. Aside from the original chunking demonstrations for verbal materials that grouped according to conceptual type (Bower & Springston, 1970), there is a growing body of evidence to suggest close synergies between short-term retention of verbal items and long-term knowledge. For example, consider the recent work of Baddeley, Hitch and Allen (2009). In their first experiment, on every trial, each participant was presented with a spoken list of words and the task was to report back the words in their presentation order. Across two conditions the words lists comprised either random words or words arranged in a sentence-like phrase (e.g., *tall solder follows waiter and old sad teacher*). Under these kinds of testing conditions a clear sentence superiority effect emerged such that words from the sentence-like lists were better remembered than were words from the random lists.

More elaborate experiments were then carried out, but the central claim was that the formation of verbal groups was, essentially, an automatic process that operates rapidly. This form of verbal grouping was taken to reflect interpretative processes concerned with the recovery of sentence structures and meaning (Baddeley et al., 2009; p. 453). It remains an open question as to exactly what is responsible for these kinds of effects. Baddeley et al. (2009) discussed several possibilities that invoke different forms of temporary memory systems that are either sensitive to the influence of long-term knowledge (i.e., their own episodic buffer) or are actually special-purpose conceptual stores (i.e., the conceptual short-term store CSTM, see e.g., Potter, 1993). Nonetheless, there is now quite a convincing case to made that the short-term retention of verbal materials can be determined by principles of organization that reflect long-term knowledge.

Although corresponding principles of organization of non-verbal materials cannot be ruled out on the basis of the present data, it appears that shared categorical membership does not provide an effective means of grouping tbr items over the short-term (see also Wong et al., 2008). This is of some interest given that there is good evidence that organization in VSTM can be determined by perceptual principles of grouping (Jiang et al., 2000; Kahneman & Henik, 1977; Quinlan & Cohen, 2012; Woodman et al., 2003). Nonetheless, it would disingenuous not to acknowledge that the case for quite separate verbal and visual short-term stores has been clouded by various demonstrations that conflicts can arise when concurrent demands are placed on remembering both visual and verbal material (see e.g., Cowan & Morey, 2007). There are many complex issues here and what appears to be key is how best to understand the synergies between attentional and memorial processes. An in-depth analysis of such issues is beyond the scope of the present work. What we are most concerned about are putative differences in how long-term knowledge affects memory for verbal and visual material, respectively. What we have failed to demonstrate here is any influence that long-term conceptual knowledge exerts on the organization of visual material over the short-term. The issue of why verbal and visual short-term memory processes appear to be differentially sensitive to the influence of properties of LTM might usefully form the subject of future research.

 It is important to acknowledge that the current findings arose in the context of single-item report experiments. It is possible that different patterns of performance may arise when whole-report tasks are used (see e.g., Vogel, Woodman & Luck, 2001). One reason why whole-report was not used in the current experiments was that in earlier pilot work participants performed very poorly. Nonetheless, in a recent innovative study on VSTM, Adam, Mance, Fukuda and Vogel (2015) described a recall technique in which memory for brief displays containing colored squares was measured. At test, the color of each tbr was probed by providing a matrix of nine possible colors at each of the item positions. Participants were instructed to choose which of the nine colors appeared at that each location. One advantage of this technique is that it allows for memory of all of the items to be tested instead of just one and it may be that under these testing conditions evidence for item grouping does emerge. This kind of experimental technique could form the fruitful basis of future work into the conceptual and perceptual factors that may influence visual memory over the short-term.

 In conclusion, in none of the experiments reported here did we find any evidence of the influence of conceptual knowledge on the memory of images over brief intervals. Our data pose a challenge to models that suggest VSTM reflects nothing other than transient activation of LTM (cf. Cowan, 1995). Such models predict category effects that we repeatedly failed to demonstrate. In contrast, our data fit more comfortably with the idea of a separate short-term visual memory system that is functionally independent of long-term memory and is pre-categorical in nature.

References

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working Mmemory capacity. *Journal of Cognitive Neuroscience, 27*, 1601-1616. doi: <http://dx.doi.org/10.1162/jocn_a_00811>

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol. 2* (pp. 89-195). New York, NY: Academic.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423. doi: [http://dx.doi.org/10.1016/S1364-6613(00)01538-2](http://dx.doi.org/10.1016/S1364-6613%2800%2901538-2)

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *Recent advances in learning and motivation, Vol. VIII* (pp. 47-90). New York, NY: Academic.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language, 61*, 438-456. doi: <http://dx.doi.org/10.1016/j.jml.2009.05.004>

Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology, 83*, 421-430.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*, 487-502. doi: <http://dx.doi.org/10.1037/a0016797>

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*, 1-34. doi: <http://dx.doi.org/10.1167/11.5.4>

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

Broadbent, D. E. (1958). *Perception and communication*. Oxford, England: Persimmon Press.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford, England: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185. doi: <http://dx.doi.org/10.1017/S0140525X01003922>

Cowan, N., & Corey, C. C. (2007). How can dual-task working memory retention limits be investigated? *Psychological Science, 18*, 686-688. doi: <http://dx.doi.org/10.1111/j.1467-9280.2007.01960.x>

Crowder, R. G. (1989). Modularity and dissociations in memory systems. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness. Essays in honor of Endel Tulving* (pp. 271-294). Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory and Cognition, 21*, 142-145.

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143*, 548-565. doi: <http://dx.doi.org/10.1037/a0033934>

Franz, V. H., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin and Review, 19*, 395-404.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5*, 351-360.

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review, 16*, 80-87. doi: <http://dx.doi.org/10.3758/PBR.16.1.80>

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects in short-term memory tasks: Evidence from a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1217-1232. doi: <http://dx.doi.org/10.1037/0278-7393.23.5.1217>

Hurlstone, M., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*, 339-373. doi: <http://dx.doi.org/10.1037/a0034221>

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 683-702.

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

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General, 139*, 558-578. doi: <http://dx.doi.org/10.1037/a0019165>

Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in perception and memory. *Proceedings of the 29th Annual Cognitive Science Society, Austin, TX,*, 407-413.

Lin, P. H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition, 17*, 356-372. doi: <http://dx.doi.org/10.1080/13506280701766313>

Lin, P. H., & Luck, S. J. (2012). Proactive interference does not meaningfully distort visual working memory capacity estimates in the canonical change detection task. *Frontiers in Psychology, 3*(42). doi: <http://dx.doi.org/10.3389/fpsyg.2012.00042>

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75*, 55-74.

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

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1-12.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Morey, C. C., Cong, Y., Zheng, Y., Price, M., & Morey, R. D. (2015). The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory. *Archives of Scientific Psychology, 3*, 18-29. doi: <http://dx.doi.org/10.1037/arc0000014>

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*, 53-81.

Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review, 20*, 1282-1289. doi: <http://dx.doi.org/10.3758/s13423-013-0460-x>

Poirer, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48A*, 384-404. doi: <http://dx.doi.org/10.1080/14640749508401396>

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43*, 489-499. doi: <http://dx.doi.org/10.3758/s13421-014-0470-6>

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509-522. doi: <http://dx.doi.org/10.1037/0278-7393.2.5.509>

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21*, 156-161. doi: <http://dx.doi.org/10.3758/BF03202727>

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10-15.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1163-1175. doi: <http://dx.doi.org/10.1037/0096-1523.28.5.1163>

Quinlan, P. T., & Cohen, D. J. (2012). Grouping and binding in visual short-term memory. *Journal of Experimental Psychology: Human Learning and Memory, 38*, 1432-1438. doi: <http://dx.doi.org/10.1037/a0027866>

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

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory, 8*, 71-78. doi: <http://dx.doi.org/10.1080/096582100387623>

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192-233. doi: <http://dx.doi.org/10.1037/0096-3445.104.3.192>

Rouder, J. N., Speckman, P. L., Sun, D., & Morey, R. D. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225-237.

Sanocki, T., & Sulman, N. (2011). Color relations increase the capacity of visual short-term memory *Perception, 40*, 635-648. doi: <http://dx.doi.org/10.1068/p6655>

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

van Lamsweerde, A. E., Beck, M. R., & Johnson, J. S. (in press). Visual working memory organization is subject to top-down control. *Psychonomic Bulletin & Review*.

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.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). Tokyo, Japan: McGraw-Hill Kogakusha, Ltd.

Wong, J. H., Peterson, M. S., & Thompson, J. C. (2008). Visual working memory capacity for objects from different categories: A face-specific maintenance effect. *Cognition, 108*, 719-731. doi: <http://dx.doi.org/10.1016/j.cognition.2008.06.006>

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10*, 80-87.

Appendix A

*Experiment 1a*

All aspects of Experiment 1 were adopted in Experiment 1a the only differences across the experiments were that a new sample of 20 University of York participants was tested and critically some of the display conditions were changed. The aim was to examine performance when both the number and size of possible groups were varied across the displays. Now five display conditions were tested including (i) a (Aaa)bcd condition and (ii) a (aaa)Bcd condition. In each of these conditions there were three grouped items and three ungrouped images. In Condition (Aaa)(bbb) there were two different groups of three grouped items. The (Aa)(bb)(cc) and Abcdef condition were the same as in Experiment 1. The results of the experiment are shown graphically in Figure A1.

*Experiment 1b*

All aspects of Experiment 1b were as in Experiment 1a except that a new sample of 20 University of York participants was tested and critically the blank duration between offset of the memory and the display was reduced from 500 ms to100 ms. The aim was to examine whether or not the amount of time prior to the probe affected item grouping. The results of the experiment are shown graphically in Figure A2

Appendix B

Table B1

*The higher-level categories and their corresponding basic level constituents used in Experiment 2*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Food | Toys | Furniture | Clothes | Tableware | Cooking Items | Outdoor Equipment | Electrical Equipment | Office Equipment |
| bagel | armyguy | beanbagchair | boot | bowl | cheesegrater | backpack | camera | pen |
| breadloaf | barbiedoll | bed | shoe | vase | cookingpan | binoculars | cellphone | ringbinder |
| cake | bearteddy | bench | glove | pitcher | cookpot | compass | gamehandheld | stapler |
| cookie | Doll | chair | hat | cupsaucer | microwave | lantern | camcorder | tape |
| muffins | dollhouse | sofa | jacket | wineglass | tongs | flashlight | mp3player | calculator |
| pizza | toyhorse | tablesmall | necktie | beermug | grill | tent | videoGameController | fan |
| sandwich | toyrabbit | dresser | pants | coffeemug | saltpeppershake | kayak | headphone | desk |

Note. The labels are taken from the photo-object database (Konkle et al., 2010).

Table 1

*Ratings for basic level constituents for each higher-level category used in Experiments 3 and 4*

|  |  |  |
| --- | --- | --- |
| Higher-level Category | Basic-level Constituent | Proportion of Respondents |
| Furniture | tablesmall | .88 |
|  | chair | .85 |
|  | sofa | .81 |
|  | bed | .81 |
| Vehicle | car\* | .92 |
|  | motorcycle | .85 |
|  | bike | .69 |
|  | bus\* | .69 |
| Musical Instrument | piano\* | .88 |
|  | guitar | .85 |
|  | violin\* | .81 |
|  | flute\* | .77 |
| Stationary | pen | .58 |
|  | ruler\* | .58 |
|  | sharpener | .50 |
|  | stapler\* | .42 |
|  |  |  |
|  |  |  |
| Electrical Equipment | tv | .73 |
|  | laptop\* | .50 |
|  | computer\* | .50 |
|  | microwave | .35 |
| Tableware | fork\* | .92 |
|  | tableknife\* | .85 |
|  | plate\* | .77 |
|  | spoon | .73 |
| Handtool | hammer | .85 |
|  | screwdriver\* | .58 |
|  | drill\* | .54 |
|  | saw\* | .54 |
| Clothes | pants | .62 |
|  | socks | .62 |
|  | shirt\* | .62 |
|  | tshirt\* | .62 |
|  |  |  |
| Shoes | boots\* | .81 |
|  | trainers\* | .73 |
|  | sandals\* | .62 |
|  | heels\* | .62 |
|  |  |  |
|  |  |  |
| Toys | doll | .54 |
|  | bearteddy | .50 |
|  | lego\* | .46 |
|  | barbiedoll | .35 |
| Food | chocolate\* | .50 |
|  | pasta\* | .38 |
|  | breadloaf | .35 |
|  | pizza | .35 |
| Cookware | spatula\* | .42 |
|  | cookingpan | .42 |
|  | saucepan\* | .35 |
|  | cookpot | .35 |
|  |  |  |

Note. Proportions are scored from the number of respondents (out of a possible 26) who included the corresponding instance in their top four responses to each higher-level category. An asterisk identifies basic level instances that were added to the original database of photo-images.

Table 2.

*Mean d’ and c scores for the four conditions of interest in Experiment 5*

|  |  |  |
| --- | --- | --- |
| Condition | *d*’  | *c* |
| Random (0) | 0.90 | 0.40 |
| Group 2 Same-In (2) | 0.81 | 0.32 |
| Group 4 Same-In (4) | 0.83 | 0.18 |
| Category (6) | 0.80 | 0.20 |
|  |  |  |

Note. Numbers in brackets refer to the number of category items in the displays.



*Figure 1.* Schematic representation of the sequence of events on a trial in Experiment 1. This an example of an (Aa)bcde trial (not drawn to scale).



a



b

*Figure 2.* a. Average *d’* scores for the display conditions of interest in Experiment 1. Error bars are simple standard errors computed for each condition. b. Mean pair-wise differences for the conditions of interest. Error bars show the 95% confidence intervals for each pair-wise difference as suggested by Franz and Loftus (2012). C1 – C2 map onto the ordered conditions in a. All CIs include 0 and hence none of the pair-wise differences are statistically reliable.



*Figure 3.* Schematic representation of the sequence of events on a trial in Experiment 2 (not drawn to scale). This is an example of an “Outdoor Equipment” trial.



*Figure 4.* Average proportion correct scores for the display conditions of interest in Experiment 2. Error bars are simple standard errors computed for each condition.



*Figure 5.* Schematic representation of the sequence of events on a trial in Experiment 2 (not drawn to scale). This is an example in which the category items are instances of “Outdoor Equipment”.



*Figure 6.* Average proportion correct scores for the display conditions of interest in Experiment 3. Error bars are simple standard errors computed for each condition.



a b

*Figure 7.* a Average d’s for the same-in and same out conditions in Experiment 3. b Average c scores for the same-in and same-out conditions. Error bars are simple standard errors computed for each condition.



a



b

*Figure 8.* a Average d’s for the same-in and same out conditions in Experiment 4. b Average c scores for the same-in and same-out conditions. Error bars are simple standard errors computed for each condition.



a



b

*Figure 9.* a Average d’s for the conditions of interest in Experiment 5. b Average c scores. Error bars are simple standard errors computed for each condition. 2in – Same-in condition for displays containing 2 grouped items, 2out – 2out – Same-out condition for displays containing 2 grouped items, 4in – Same-in condition for displays containing 4 grouped items, 4out – Same-in condition for displays containing 4 grouped items.



*Figure 10.* Average proportion correct scores for the display conditions of interest in Experiment 6. Error bars are simple standard errors computed for each condition.



a b

*Figure 11.* a Average *d’*s for the same-in and same out conditions in Experiment 6. b Average *c* scores for the same-in and same-out conditions. Error bars are simple standard errors computed for each condition.



a



b

*Figure A1.* a. Average *d’* scores for the display conditions of interest in Experiment 1a. Error bars are simple standard errors computed for each condition. b. Mean pair-wise differences for the conditions of interest. Error bars show the 95% confidence intervals for each pair-wise difference as suggested by Franz and Loftus (2012). C1 – C2 map onto the ordered conditions in a. All CIs include 0 and hence none of the pair-wise differences are statistically reliable.



a



b

*Figure A2.* a. Average *d’* scores for the display conditions of interest in Experiment 1b. Error bars are simple standard errors computed for each condition. b. Mean pair-wise differences for the conditions of interest. Error bars show the 95% confidence intervals for each pair-wise difference as suggested by Franz and Loftus (2012). C1 – C2 map onto the ordered conditions in a. All CIs include 0 and hence none of the pair-wise differences are statistically reliable.