Intention, Attention and Long-term Memory for Visual Scenes: It all depends on the scenes

Karla K. Evans & Alan Baddeley
University of York

Authors:
Corresponding author:

Karla K. Evans

Department of Psychology

The University of York

York, North Yorkshire, UK

Email: karla.evans@york.ac.uk

Tel. +44-01904-32-4601

Fax. +44-01904-32-3190

Alan Baddeley

Department of Psychology

The University of York

York, North Yorkshire, UK

Email: a.baddeley@york.ac.uk

Tel. +44-01904-32-2882

Fax. +44-01904-32-3190

Running Head: ?????

Abstract

Humans have an ability to remember up to 10,000 previously viewed scenes with apparently robust memory for visual detail, a phenomenon that has been interpreted as suggesting a visual memory system of massive capacity. Attempts at explanation have largely focused on the nature of the stimuli and been influenced by theoretical accounts of object recognition. Our own study aims to supplement this by considering two observer-based aspects of visual long-term memory, one strategic, whether the observers are aware or not that their memory will subsequently be tested and the other executive, based on the amount of attentional capacity available during encoding. We describe six studies involving visual scenes ranging in difficulty from complex manmade scenes (d' = 2.54), to door scenes with prominent features removed (d' = 0.79). To ensure processing of the stimuli, all participants have to make a judgement of pleasantness (Experiments 1 and 2) or of the presence or absence of a dot (Experiment 3). Intention to learn influence performance only in the most impoverished condition comprising doors with prominent features removed. Experiments 4 to 6 investigated the attentional demands of visual long-term memory using a concurrent task procedure. While the demanding task of counting back in threes clearly impaired performance across the range of materials, a lighter load, counting back in ones influences only the most difficult door scenes. Detailed analysis of error patterns indicated that clear differences in performance level between manmade and natural scenes and between unmodified and modified door scenes was reflected in false alarm scores not detections, while concurrent task load affected both. We suggest an interpretation in terms of a two-level process of encoding at the visual feature rather than the whole scene level, with natural images containing many features encoded richly, rapidly and without explicit intent. Only when scenes are selected from a single category and with distinctive detail minimised does memory depend on intention to remember and on the availability of substantial executive capacity.

Keywords: visual episodic memory, executive attention, intent to learn, encoding, working memory.

Introduction

The psychology of human memory has been actively pursued for over a century, resulting in a rich blend of evidence and theory that successfully links detailed analysis within the laboratory to the practicalities of everyday experience (Baddeley, Eysenck & Anderson, 2014). The area has however been heavily dominated by theories developed using memory for verbal material, typically lists of unrelated words. There are good reasons for this; the material is easily accessible, scoreable and manipulable and has generalised readily to more complex verbal material such as sentences and prose, when factors such as syntax and semantics become important. The world is not however made up entirely of words, and there is clear evidence from neuropsychology that visual and verbal memory may be differentially disrupted (Vallar & Shallice, 1990; De Renzi & Nichelli, 1975).

The study of visual long-term memory (LTM) has however, until recently, been somewhat limited. In the clinical domain, it has tended to rely on memory for abstract designs such as the Rey figure (Rey, 1964) or recognition memory for unfamiliar faces (Warrington, 1984). Neither of these is representative of the visual world more generally; the Rey figure introduces complexities from the motor and constructional demands of drawing, while faces, although very important ecologically are not typical of the rest of the visual world, having a strong social connotation with links to emotion and to processing in specialised brain areas (Öhman, 2009).

An exception to this comparative neglect is reflected in the recent rekindling of interest in the dramatic demonstration of the apparently huge capacity of visual LTM (Standing, Conezio & Haber, 1970; Standing, 1973). Participants shown up to 10,000 pictures selected from magazines were able to recognise those seen previously with an 83% accuracy (Standing, 1973). More recently, Brady, Konkle, Alvarez and Oliva (2008) raised the issue of whether performance could be based on the general gist of the overall scene, rather than on retention of specific detail. They selected objects from a wide range of conceptual categories, demonstrating recognition rates of 88% even when remembered items and foils came from the same category, concluding that performance could not be explained on the basis of gist retention. A subsequent study by Konkle Brady, Alvarez and Oliva (2010a) investigated the role of semantic category membership in more detail, presenting 2800 pictures of objects. The number of exemplars per category ranged from 1-16, showing a systematic though modest increase in error rate with increase in exemplar set size, 2% for every doubling of number of category items within the list. However, most categories involved nameable objects, possibly implicating an additional verbal component. It is notable in this regard that Paivio (1971) using line drawings of nameable objects showed that subsequent retention was substantially better for nameable drawings than was memory for the names alone. He interpreted this in terms of a dual code hypothesis whereby the additional information provided by parallel verbal and visual codes reduced forgetting. A further study (Konkle, Brady, Alvarez & Oliva, 2010b) minimised the potential contribution of verbal labelling by using a limited range of scenes either natural or man-made. These were much less individually nameable than objects, and yielded broadly equivalent results, although at a somewhat lower level of performance.

The theoretical drive behind these recent developments has come predominantly from the study of vision. Brady et al. (2008) applied information theory to estimate the capacity of visual LTM, and used theories of object recognition as a basis for possible explanations. Another potentially fruitful theoretical link is provided by work on computer-based scene analysis. Isola Xiao, Parikh, Torralba and Oliva (2014) used this approach to determine what makes a picture memorable. They found highly consistent differences in the memorability of pictures, differences that, somewhat surprisingly were not related to subjective ratings of memorability. A range of purely visual variables proved equally ineffective, although performance improved when these were combined with subject-based reports on features of the various pictures suggesting for example that pictures of people were typically more memorable than objects, which in turn were better than natural scenes.

This, together with the importance of semantic category membership shown by Konkle et al. (2010a) suggests a role for semantics in visual memory. In this respect, visual memory resembles verbal memory where the crucial feature determining memorability is not the word itself, but the semantic representation that it generates. If the word bank is initially presented in the context of money, it is much less likely to be recognised when the context is switched to a river (Light & Carter-Sobell, 1970). However, although there are broad similarities between what we know of visual and verbal LTM, it seems likely that there will also be differences. For example, a recent study by Baddeley and Hitch (2017) found that the Levels of Processing effect (Craik & Lockhart 1972), whereby deeper and more elaborative processing of verbal materials leads to better retention, is potentially much more powerful for verbal than for visual material, a result they interpret in terms of the reliance of verbal memory on potential for semantic elaboration. This they suggest will depend on the encoding instructions together with the semantic richness of the stimuli. In contrast they suggest that visual features tend to be rapidly and richly encoded but typically have less potential for further semantic elaboration. They go on to interpret their results in terms of Nairne's feature model of long-term memory (Nairne, 2002). The experiments in the current paper reflect a further exploratory attempt to supplement earlier explanations of the apparently massive capacity of visual LTM by investigating the processes of encoding focusing on two aspects, one strategic through the effects of intention to learn and the other using a concurrent task methodology to study the importance of attentional demand. We study both effects across a range of stimuli varying widely in visual characteristics and memorability.

Encoding can be intentional with observers potentially applying organizational strategies to the material at hand, or incidental with no intention to remember. In the case of verbal material, amount retained depends crucially on the encoding task, regardless of whether or not there is intention to remember (Craik & Lockhart 1972; Mandler 1967) while the nature of the encoding task appears to be much less critical for visual material (Castelhano & Henderson, 2005; Baddeley & Hitch 2017).

While elaborative coding of verbal material is advantageous, it is also attention-demanding. Direct evidence of the importance of executive resources in remembering comes from studies using concurrent secondary tasks. Free recall of lists of unrelated words was disrupted by a concurrent card sorting task with the degree of disruption depending on the concurrent information load as varied by number of sorting alternatives (Murdock, 1960; Baddeley, Scott, Drynan & Smith, 1969). Verbal paired-associate learning also shows substantial disruption from concurrent attentional demands (Baddeley, Lewis, Eldridge & Thomson, 1984), as does prose recall (Baddeley & Hitch, 1974). There have been fewer studies that have examined memory for visuo-spatial material with the general pattern of findings from a dual-task manipulation broadly like that reported for verbal materials (Fernandes & Guild, 2009). But we know of no equivalent research on memory for the visually complex and semantically diverse scenes that are typically used in "massive memory" visual recognition studies. However, given that performance remains at a very high level in the classic studies, despite up to 10,000 pictures being presented over a period of many hours, it seems unlikely that effortful elaborative processing would be sustained for this length of time, suggesting that the encoding of visual material in long-term memory may require fewer resources, be more resilient to load and require less intent.

Across the 6 experiments we examine how the two aspects of encoding (intent to learn and availability of attention) interacts with the nature of the material to be encoded, (visual and semantic). Our first three experiments investigate the role of intention to remember and in the last three the availability of attention across a range of stimulus types varying in visual and semantic complexity using both deep and shallow encoding tasks. We compare scenes with manmade structures to scenes of nature in Experiments 1, 2 and 4, moving on in Experiments 3, 5 and 6 to a single broad category, door scenes, both complete and modified by removing distinctive detail (Vogt & Magnussen, 2007). We finally combine the results of our two approaches to allow general conclusions to be drawn.

Intent to Learn and Encoding

One way to probe the effects of encoding method is to test if it requires intent. Thus, in Experiments 1 to 3 we studied intention to remember using the standard massive memory paradigm. We further manipulated the visual material on dimensions of both category-based semantic and visual complexity. Observers were presented with pictures of 400 complex scenes, (Experiment 1 & 2) and 304 door scenes, (Experiment 3). In each experiment, observers were randomly assigned to two conditions, intentional or incidental, where only the observers in the intentional memory group were aware that their memory for images would be subsequently tested. We then combined data from the three experiments to yield an overview of our results.

All experiments used yes/no decision, with items to be recognised randomly mixed with an equal number of “new” foils. While this differs from the two-alternative forced choice method used in some of the earlier studies, recent studies involving both stimuli from a large database of door scenes (Baddeley, Hitch, Quinlan, Bowes & Stone, 2016) and complex real world scenes (Evans, Cohen, Tambouret et al., 2010) suggests broadly similar d’ scores across studies using yes/no, two-alternative and four-alternative forced choice.

Experiment 1

In the first experiment, we look at the interaction of intent to encode with visual material that is varied on both the sematic and the visual dimension. We manipulate the intent to encode by using a separate groups design in which all participants were required to make a pleasantness judgment on all stimuli but only one group was informed that a memory test of the scenes would follow. We selected pleasantness as it has been shown to be an effective means of ensuring overall processing of stimuli resulting in a consistent but modest enhancement in subsequent recognition of visual materials when compared to instruction to judge a single visual feature (Baddeley & Hitch, 2017).

Subjects

A total of 28 participants (18 female, aged 18-30) gave informed consent, with 14 participants randomly allocated to one of the two conditions. The sample size was based on a power analysis of previous published work (Evans et al., 2010) using the same paradigm, yielding an estimated effect size of d = 1.74 (Cohen's d). An effect size of 1.74 α = .05 and power = 0.95 returns a minimum sample size of 7. The chosen sample size of 14 has a power of 0.99 with d = 1.74, α = .05. All participants had normal color vision as tested by the Ishihara test, normal or correct vision and all were York student volunteers rewarded either by course credit or by a modest payment.

Stimuli and Apparatus

There were 400 images comprising 100 exemplars from each of four categories (interiors, cityscapes, waterscapes and landscapes) obtained from a public image dataset hosted by the Computational Visual Cognition Laboratory MIT (http://cvcl.mit.edu). We use two broad semantic categories of images with manmade structures (e.g. interiors and cityscapes) typically relatively rich in complex cues and images of natural scenes with less complex detail typically involving relatively uninformative redundancy (e.g. waterscapes and landscapes). Examples are shown in Figure 1a. The scene stimuli subtended 13° by 13°of visual angle.

Participants were seated approximately 57cm from the monitor and stimuli presented on 21 in CRT monitor set to a resolution of 1280 by 1024 at a refresh rate of 85Hz, controlled by a Dell

XPS computer running Windows 7. The experiment was controlled by Matlab R2013 and the Psychophysics Toolbox version 3 (Brainard, 1997; Pelli, 1997).

Figure 1

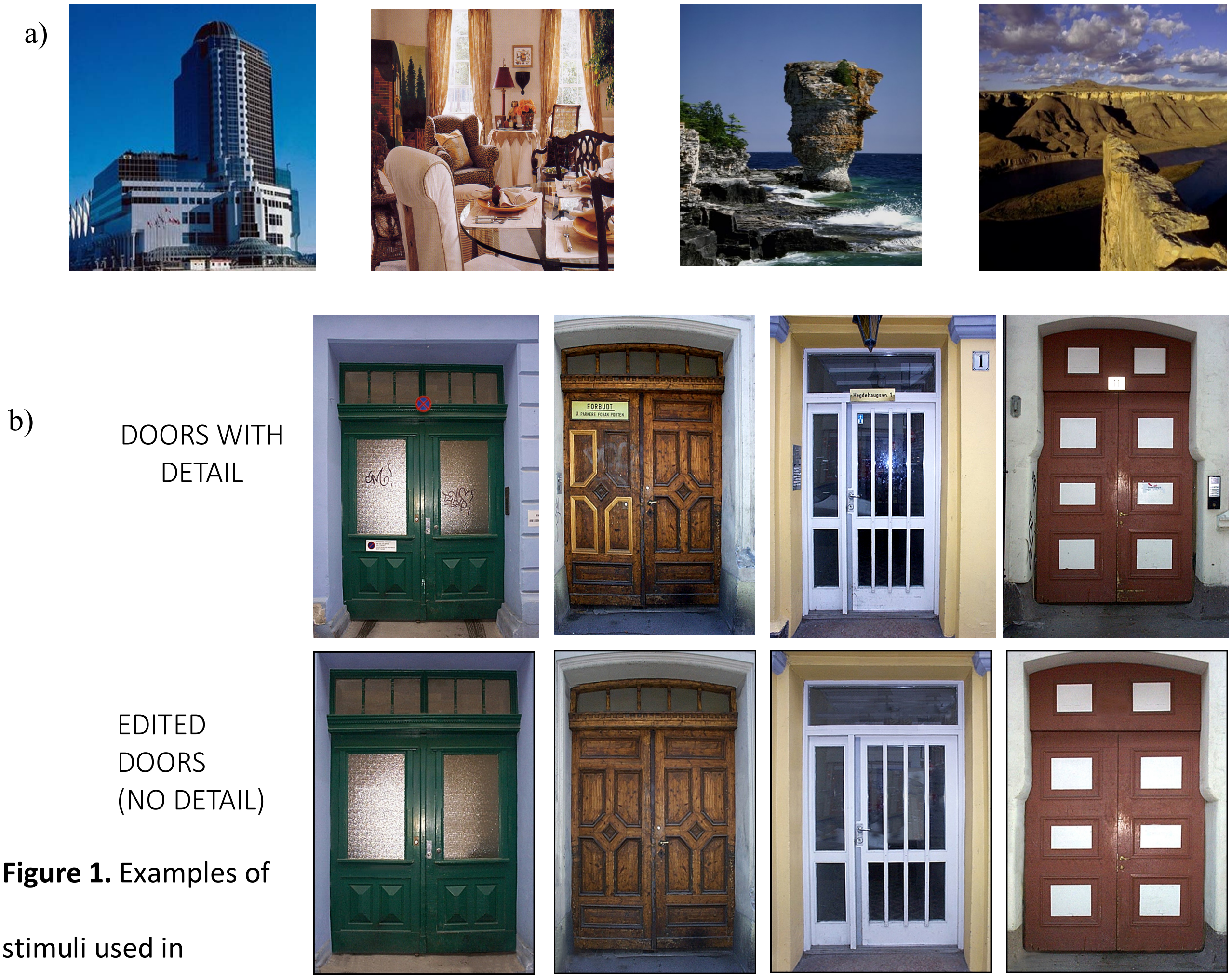


Figure 1. Examples of stimuli used in different experiments: a) scenes form 4 different categories (cityscapes, interiors, waterscapes and landscapes) used in Experiment 1,2 & 4; b) scenes of doors (original and edited) used in Experiment 3,5 & 6.

Procedure

The experiment comprised a study phase in which 200 images were presented, followed by a test phase in which a 100 of the previously judged images were randomly intermixed with 100 completely new images from the same four categories as in the study phase for an Old/New recognition task. During phase 1, each image was presented for 3s with observers required to judge its pleasantness on a four-point scale, a standard method of encoding that ensures a degree of attention to each stimulus in both intentional and incidental groups (Baddeley & Hitch, 2017; Warrington, 1984). The recognition test phase was self-paced with images presented singly and remaining until the participant responded. Participants responded by pressing an "old" or "new" response key. Feedback was provided for each test image. Performance in this experiment and all subsequent ones was measured in terms of d' after correcting the cells with perfect performance by adding 0.5 to the frequency of hits and false alarms and dividing by N + 1 where N is the number of old or new trials (Snodgrass & Corwin, 1988). False alarms and correct rejections were binned separately for each stimulus type in this and all subsequent experiments.

Results

We manipulated two factors in this experiment, the intent to learn (incidental versus intentional encoding condition) and the complexity of the encoded material, manmade versus natural scenes (see Figure 2a). A two way mixed model ANOVA showed a significant main effect of image type ($F(1,26) = 100.53$, $p < .001$), with manmade structures comprising cityscapes and indoor scenes being better remembered (d' = 2.54, s.e.m = .15) than natural scenes (d' = 1.45, s.e.m = .11). However, the overall effect of encoding instruction failed to reach significance ($F(1,26) = 0.017$, $p = .90$), as did the interaction between instruction and material ($F(1,26) = 1.18$, $p = .29$). There was no difference between level of memory performance in participants who were instructed to remember (d' prime = 1.85, s.e.m = .13) and those who were not expecting a memory test (d' prime = 1.85, s.e.m = .19), consistent with the assumption that the encoding of such scenes does not require intent.

Figure 2

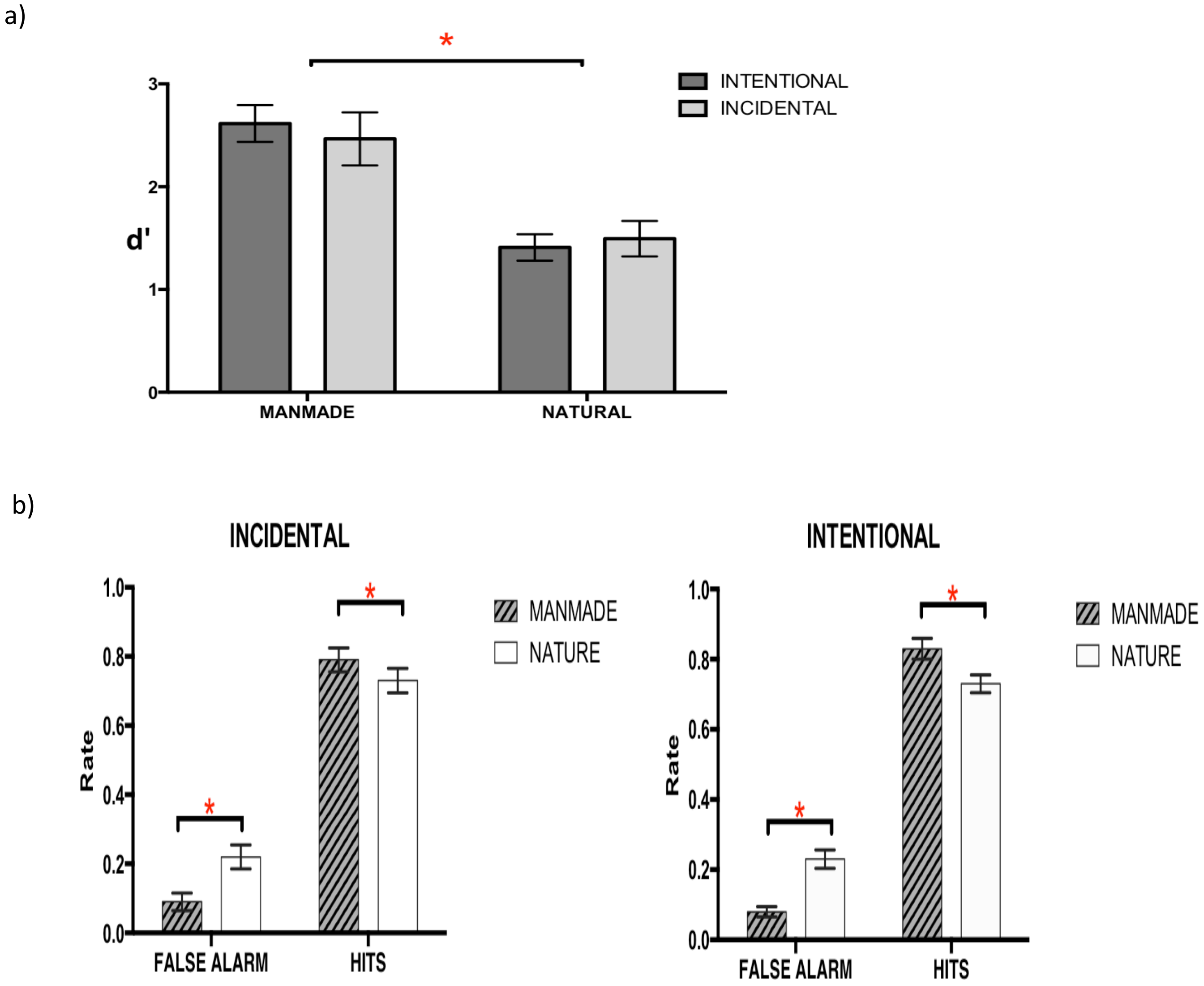


Figure 2. a) Performance on visual recognition memory by task type (intentional vs. incidental) and image type (manmade vs. natural scenes). b) Histograms of False alarm and Hit rates separated by task type and image type. *p<.05

Finding a significantly better recognition performance for manmade scene images than natural images led us to explore further the origin of the memory advantage, in particular, whether the observers are more likely to correctly recognizing seeing an old image, a hit response (HIT) or are less likely to incorrectly recognize a new image as seen before, false alarms response (FA) for each type of image. For this we conducted two separate mixed model ANOVAs with the intent to learn (incidental vs. intentional) as a between group factor and the image type factor (manmade vs.

natural) as within factor separately for the two kinds of response, hits and for false alarms. The main effect of type of image was significant (F(1,26)=65.07,p<.00001; see Figure 2b) in false alarm rates with a higher false alarm rate (22%) for natural than for manmade scene images (8%). The performance difference between two image types was also statistically significant for the hit rate (F(1,26) =14.60,p=.001), but they differed to a lesser degree (hit rate for natural 73% for manmade scene 81%). The effect of intent to learn was not evident in the false alarm rates (F(1,26) =.71, p=.407) nor in the hit rate (F(1,26) =.47, p=.501) . There were also no interactions between image type and the intent to learn in either of the response types (FA p=.485; HIT p=.354).

Discussion

Experiment 1 failed to produce evidence that intention to learn enhances performance, provided participants attend to the relevant scene. However, we see that observers are both more likely to have a higher hit rate and a lower false alarms rate for manmade than for natural scene images during recognition.

While a negative result of intent to learn can always be attributed to lack of power, Figure 2 gives no support to such an interpretation. It could however be suggested that our judgement task itself induces a very effective learning strategy, indicating the need to replicate using a task that requires scanning the stimuli but processing at a more superficial level, hence likely to lead to lower overall performance. For this purpose, we adopted a task requiring detection of one or more dots that could occur anywhere on the scene.

Experiment 2

Subjects and Design

In Experiment 2 we tested another group of 28 participants (17 female, aged 18-30) with one group of 14 participants not aware and another group of 14 aware that their memory would be tested. The design was similar to Experiment 1 except that we used an orthogonal task instead of pleasantness rating, asking participants to detect an appearance of a dot anywhere on the displayed image and indicate with a button press as to whether the dot was white or grey and then further monitor for a possible appearance of a second dot. The first dot was present from the start of the image presentation while the appearance of the second dot was jittered in time and did not appear on each of the images to remember enticing the observers to continue monitoring each image for the full 3 seconds of image presentation. The image by image time restricted presentation and a dot detection and discrimination was followed by an old/new recognition phase that was self-paced. In selecting the dot detection and discrimination task our aim was to further restrict the possibility for the incidental condition group to be able to engage in explicit coding. The stimuli that were to be encoded were the same as in Experiment 1 with two manmade image categories and two natural image categories.

Results

Data analysis was, as in Experiment 1, a two-factor mixed model ANOVA with encoding condition (between group) and image type (within group) factors. Even with the new orthogonal task there was no significant difference (F(1,26)=.105, p=.601) between performance of intentional encoders (d'=.67, s.e.m=.15) and incidental learners (d'=.61, s.e.m=.14). Manmade scenes (d'=.77, s.e.m=.17) were again remembered better (F(1,26)=31.65, p<.0001) than natural scenes (d'=.51, s.e.m=.13) (See Figure 3a). But there was no interaction between type of encoding and type of material encoded (F(1,26)=.280, p=.601). It is important to note that the performance on the dot detection and discrimination task during the encoding phase of the experiment did not differ between the intentional (96%, s.e.m. 3%) and incidental (96%, s.e.m. 4%) encoding conditions (p=.939).

Figure 3

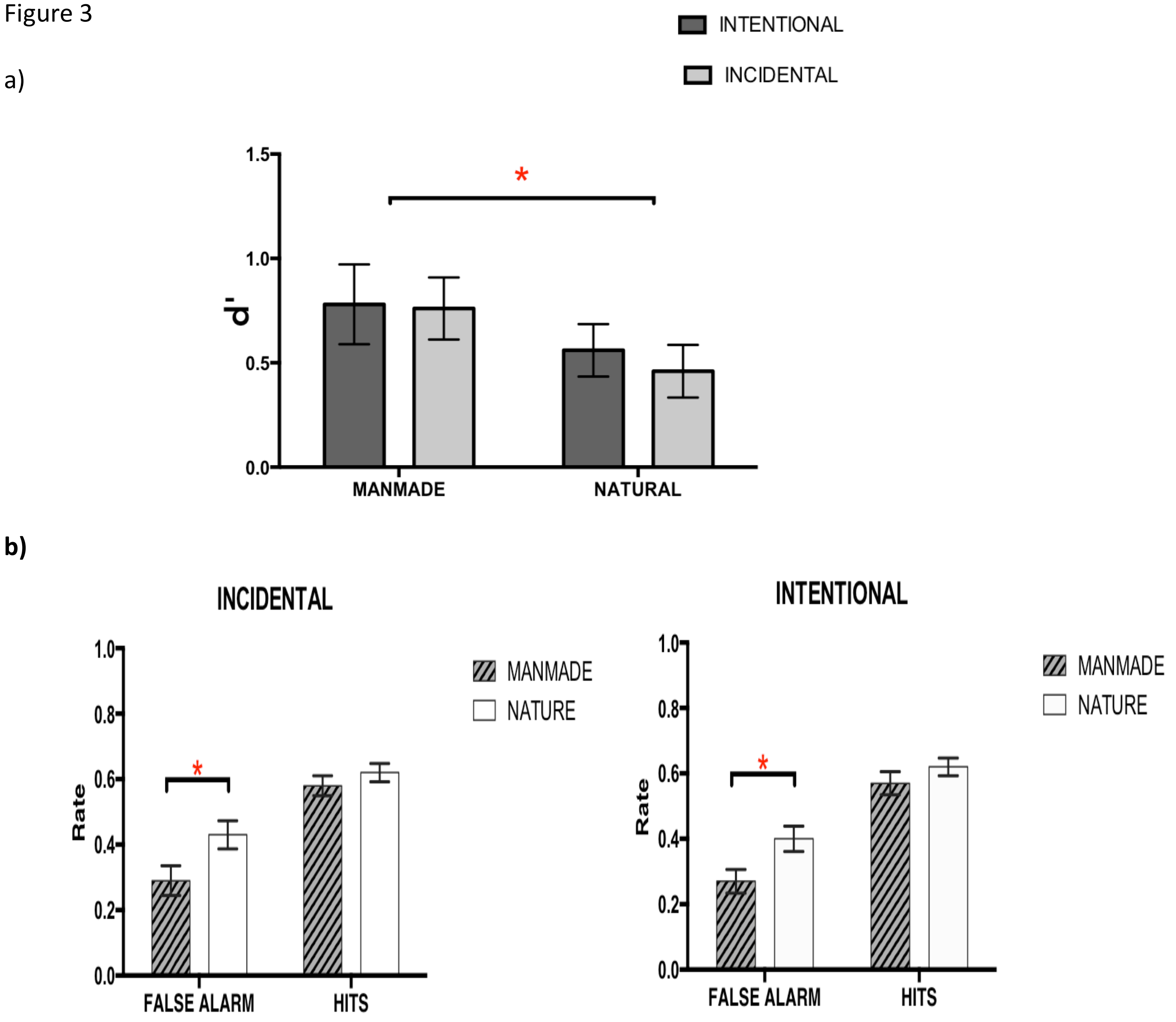


Figure 3. Performance on visual recognition memory, d' by task type (intentional vs. incidental) and image type (manmade vs. natural scenes). b) Histograms of False alarm and Hit rates separated by task type and image type. *p<.05

As in Experiment 1, we ran two mixed model ANOVAs in this experiment to investigate separately the effects of image type and encoding conditions on false alarm and hit rates. Observers in this experiment show a significantly higher false alarm rate (42%) for natural than for manmade scene images (28%) (F(1,26)=45.53, p<.0001; Figure 3b) independent of the encoding condition (F(1,26)=.273, p=.603). However, they show no significant difference in hit rates (natural 62% vs.

manmade 58%, F(1,26)=3.91, p=.06). The encoding condition did not have any main effect (F(1,26)=.002, p=.969) on the hits nor did it interact with the images type (F(1,26)=.101, p=.752).

Discussion

Experiment 2 using a demanding secondary dot detection task in comparison to pleasantness judgment failed to produce evidence that intention to learn enhances performance even though the dot task substantially reduced the overall encoding performance in both intentional and incidental condition. This is evident from the overall lower performance in Experiment 2 (d'=.64) than in Experiment 1 (d'=1.99) for the same encoding material. Observers retained visually and semantically varied visual material to the same extent whether they were encoding incidentally or intentionally. There is however a suggestion of a difference for the more difficult material that we suggest may be interpretable within a feature model (Nairne, 2002; Baddeley & Hitch, 2017). This assumes that the memory trace depends on two factors, the number of features retained and their "diagnosticity" that is the capacity of such features to distinguish the to-be-remembered item from "new" items. We suggest that manmade scenes contain a richer and more varied set of potentially memorable and diagnostic features than do our natural scenes. We consistently observe (both in Experiment 1 & 2) that for this type material there is a significantly higher false alarm rate than for material with more potentially diagnostic features.

Given that our approach has been essentially exploratory, we did not set out to test such a hypothesis. Our next experiment therefore follows this up by using material chosen explicitly to allow a comparison between conditions that vary only in the presence or absence of potentially diagnostic features. We select material from a single instead of four basic scene categories (door scenes), hence reducing semantic richness. Images from that single category comprised two conditions in one of which detailed visual features are removed, resulting in two versions of the same door, in one of which potentially diagnostic features had been minimised, for example removing a doormat, idiosyncratic letter box or light fixture. Our comparisons are thus based on the presence or absence of such features, independent of the visual complexity of the door itself. This was made possible by using door scenes, with and without the removal of such idiosyncratic features, using material provided by Vogt and Magnussen (2007) who had shown that removal of such features led to a clear drop in retention.

In selecting material for Experiment 3 we thus aimed to follow up our interpretation of the observed difference between manmade and natural scenes by comparing two sets of images from one sematic category carefully selected to differ in the presence or absence of detail of a type that might be expected to differ in the number of potentially diagnostic features.

Experiment 3

Subjects and Design

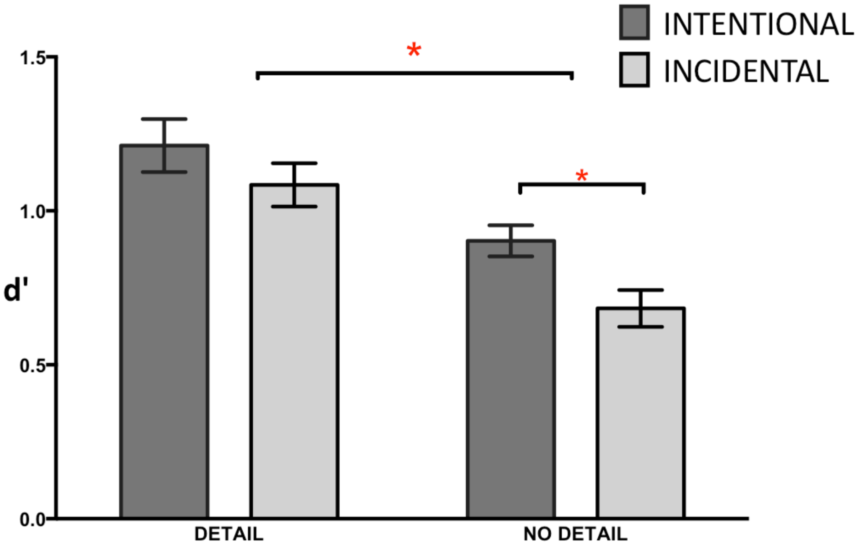
We tested a new group of 56 participants (37 female, aged 18-30) with one group of 28 participants not aware and another group of 28 aware that their memory would be tested. The design was identical to Experiment 1, with the manipulation of intent to learn (intentional and incidental) with both groups making pleasantness ratings. The stimuli comprised the color photographs of doors and their immediate surroundings used by Vogt and Magnussen (2007). Using versions of each door with or without detail in separate lists, we could test for a possible interaction between intention to remember and available detail. Fewer stimuli were available than in Experiments 1&2 hence both the study and test sets comprised 200 images, half of which had discriminative detail removed. Each image occurred in its detailed form for one group and its reduced form for the other. Thus, a door scene with a door of same complexity could be shown with all the redundant detail intact in one stimulus set or the detail stripped away in the other. The two stimulus types were intermixed in both the study and test phase. The observers were not informed of the presence of the two stimulus types. Images were again presented for 3s while a pleasantness judgement was made, and again the old/new recognition phase was self-paced.

Results

Results of Experiment 3 are shown in Figure 4a. Data analysis was as in Experiments 1&2, except that the data for original and modified images were treated as a further within-participant factor as a part of two factors mixed ANOVA. This yielded a significant main effect of intention to learn (*F*(1,46) = 4.25, *p* = .045) with intentional encoders (d'= 1.05, s.e.m = .07) performing better than incidental learners (d'= .88, s.e.m =.08). There was a substantial effect of removing the image detail (*F*(1,46) = 90.62, *p* < .001) with doors including detail yielding consistently better performance (d' = 1.15, s.e.m = .13) than those with detail removed (d' = .79, s.e.m = .11). There was also a significant interaction between instruction and material (*F*(1,46) = 5.05, *p* = .03), reflecting the fact that intention to learn failed to reach significance for images with more detail (t(46)=1.15, p=.26) whereas there was a clear effect of intention to learn for the more difficult condition with detail removed, (t(46)=2.81, p=0.007).

Figure 4

a)



b)

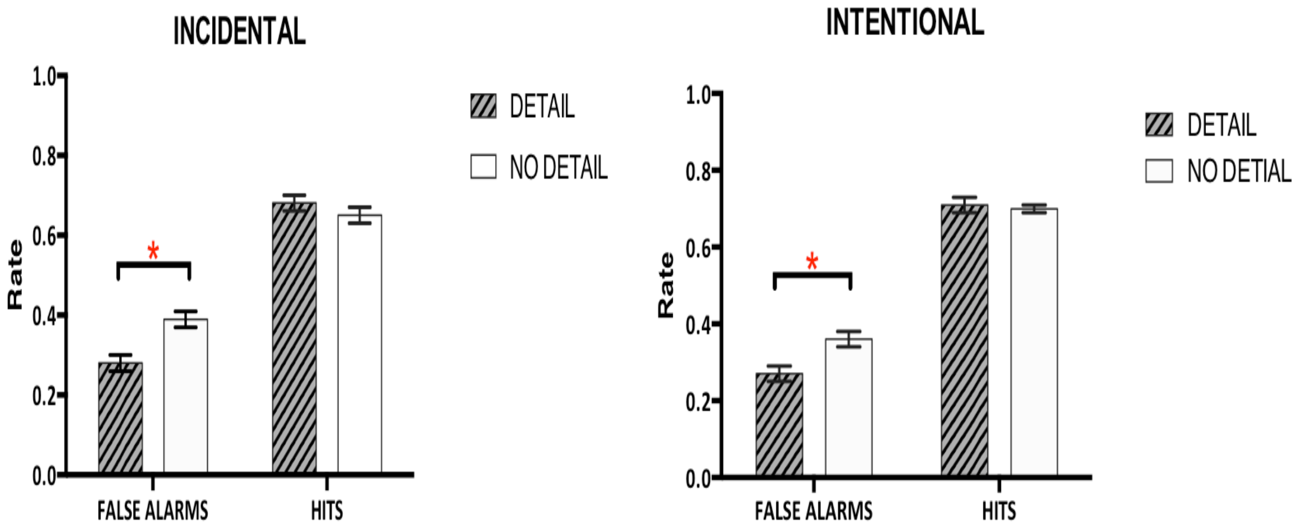


Figure 4. a) Performance on visual recognition memory in Experiment 3 by task type and image type (doors with detail vs. doors with no detail). b) Histograms of False alarm and Hit rates separated by task type and image type. *p<.05

When looking at false alarms and hits in separate ANOVA's we again observe that it is only the false alarms that show a significant effect of level of image detail (F(1,46)=50.09, p<.0001) while hits are unaffected (F(1,46)=2.27, p=.139). Observers showed a higher false alarm rate for images with less detail (37%) as opposed to images with detail (27%), whereas no such difference was evident in the hit rates for the two types of material (images with detail 70% versus without detail 68%, Figure 4b). The main effect of intent to learn was not evident for either response types (FA p=.440; HIT p=.522) nor did it interact with the level of image detail (FA p=.475; HIT p=.491).

While choice of hedonic judgement used in Experiment 1 & 3 was taken principally because it is likely to involve processing the picture as a whole, it does incidentally allow two further questions to be asked. The first of these concerns whether memory is better for scenes that are judged as more pleasant, a possibility that is raised by the observation by Isola et al. (2014) who found that aesthetic judgements correlate positively with subjective judgments but negatively with objective measure of memorability across a wide range of images, perhaps surprisingly suggesting that aesthetically pleasing stimuli are less well remembered. Experiment 1 showed no difference in memorability between images rated as more or less pleasant, nor did pleasantness interact with intention to remember (*F*(1,398) =.318, *p* = 0.57). In the case of the door stimuli however, a difference did occur, with the doors (Experiment 3) rated as more pleasant recalled somewhat better (70%, s.e.m 2%) than those rated as less pleasant (66%, s.e.m 2%) a difference that was significant (*F*(1,74) = 7.56, *p* < 0.01)), , in direct contrast to the results from Isola et al (2014) who found a negative effect of rated pleasantness. Rated pleasantness did not however interact either with amount of detail in the image or the intent to encode. It seems possible that the small positive effect of pleasantness shown in Experiment 3 may have been swamped by other factors in the more complex and richly encodable material used in Experiment 1.

Discussion

Results of Experiment 1 to 3 are broadly consistent with the assumption that the type of material that characterises the typical massive memory paradigm studies can be encoded without explicit intent to learn. Two different secondary tasks failed to differentiate memory performance between the condition in which observers were asked to explicitly encode the images and the condition in which observers had no reason to remember the images. However, intention to remember does enhance performance in the case of the most difficult material, the door scenes with distinguishing detail removed when we not only limit the semantic but also restrict the visual distinctiveness of the to be encoded material.

Figure 4 suggests that the situation changes when images are stripped of idiosyncratic detail raising the issue of just how the intention to learn might improve performance on these more difficult items. One possibility is that active intention to remember may lead to a selective choice of potentially diagnostic features, or simply to an increase in the number of features encoded, resulting in a greater probability that potentially diagnostic features might be stored. Intention to learn thus has little effect on retention of stimuli that are already richly encoded since there are already enough visual features to allow discrimination without the need to search for other potentially diagnostic cues.

On the assumption that operating such a strategy might be demanding of attentional capacity, we suggest that a concurrent task placing demands on working memory might show a similar pattern, having little impact on stimuli with multiple potentially diagnostic cues, while potentially reducing performance as the available discriminative cues decrease. Experiments 4 to 6 explore this hypothesis, again, using both natural scenes and door scenes.

Executive Attention and Encoding

Experiments 4 to 6 probed the role of executive attention using a within-subject design that compared three concurrent load conditions, in all of which observers were informed that recognition memory for the stimuli would be tested. In these experiments we increase the sample size from 14 to 24 observers to compensate for reduced number of trials per condition as we use the same amount of stimulus material as in Experiments 1 to 3 but increase the number of conditions.

Experiment 4

Subjects

We used a within-participant design in which each participant was tested on each of three load conditions, baseline, low load and high load in counterbalanced order. A total of 24 new participants (15 female, aged 18-30) were tested and all were informed that memory would subsequently be tested. All gave informed consent, passed the Ishihara test of color blindness, had normal or corrected vision and were rewarded with credit points or a small payment.

Stimuli and Apparatus

We again used the four categories of scenes from Experiments 1&2, namely interiors, cityscapes, waterscapes and landscapes adding additional stimuli to make an equal number of images across the three load conditions, resulting in a total of 452 distinct images with equal representation of the four categories. The scene stimuli subtended 13° X 13° of visual angle. The apparatus used in Experiment 4 was the same as in Experiments 1 to 3.

Procedure

We tested for the effect of executive function by having participants encode under different working memory load conditions based on the standard approach of backward counting (Peterson & Peterson,1959) and varying demand by manipulating the size of the counting steps. During the study phase a baseline condition was compared with a low load condition involving counting back in ones and a high load involving backward counting by threes. Images were again presented for 3s. In order to allow sufficient time for counting to be established, we blocked images in groups of five. Each condition thus involved presenting a total of 20 sets of five images, each image presented for 3s. Before each block of five, a three-digit odd number was presented in a male voice through the headphones (e.g. 285). In the baseline condition, participants were instructed to ignore the voice, in the low load condition they were instructed to count backwards in ones, while in the high load they counted backwards in threes, in each case reporting the total at the end of each block of five images seen.

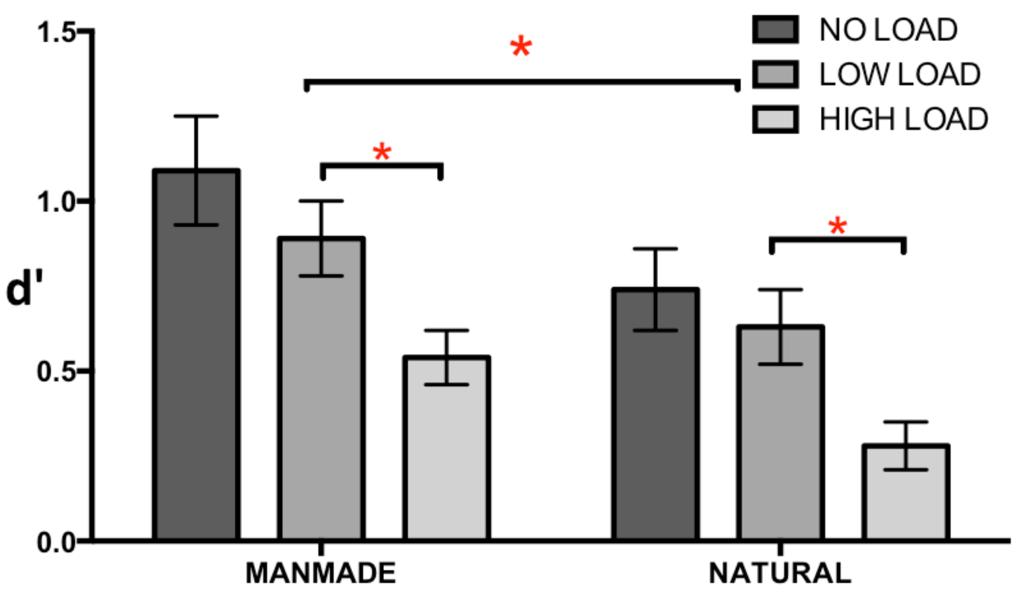
Participants were aware that their memory would be tested, and no hedonic judgement was required. Each participant was tested in each of the three conditions with order of conditions in the study phase counterbalanced using a Latin Square. In the study phase each participant saw a total of 300 images. Recognition memory was tested after all three-study phase conditions were completed using the same self-paced recognition procedure as in Experiments 1 to 3. In the test phase the participants were shown half of the images, they had seen in the study phase, randomly intermixed with an equal number of new images not previously viewed (for a total of 300 images) and they were asked to make an old/new judgment.

Results

Mean performance across the three conditions and two image types is shown in Figure 5a. A repeated measures two-factor ANOVA indicated a significant effect of load (F(1.32, 30.34)= 17.55, p<.001). Tests of within-subject planned contrast analysis indicated that counting backwards in ones had a marginal effect on performance (F(1, 23)= 3.94, p=.059, two-tail), while performance declined significantly when counting back in threes (F(1, 23)= 30.43, p<.001). We again find a reliable difference between two supra-category image types (*F(*1,23) = 17.55, *p* <.001) with overall better performance for manmade environments and no interaction between image type and load (F(1.52, 34.88)= 1.35, p=.27).

Figure 5

a)



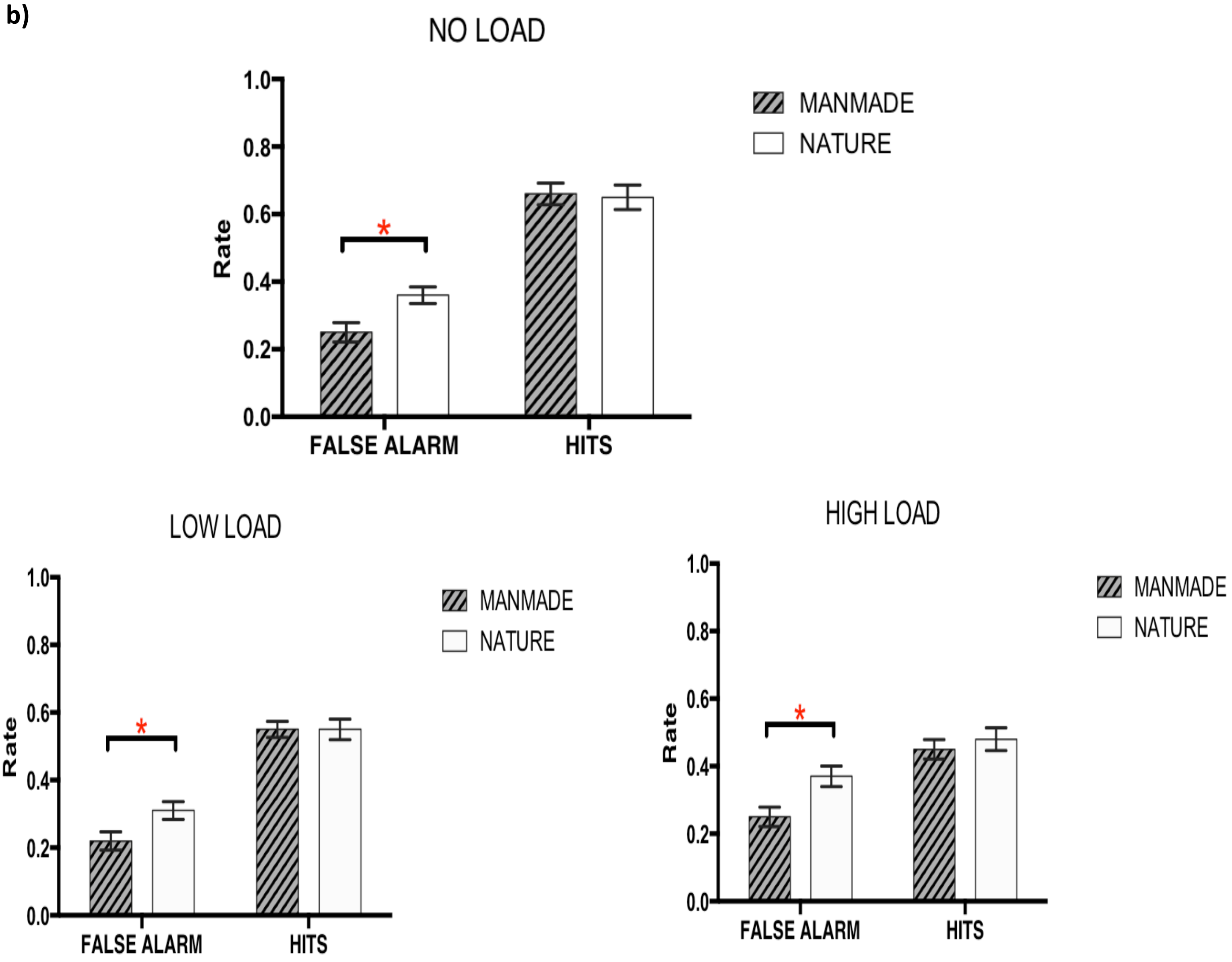


Figure 5. a) Performance on visual recognition memory by load and image type. Data from Experiment 4 across different image types (manmade vs. natural scenes) and across different loads. b) Histograms of False alarm and Hit rates separated by load and image type. *p<.05

Here too, we ran two mixed model ANOVAs separately for FA and HITs to investigate the origin of the material type advantage and concurrent load effect. From graphs in Figure 5b it is evident that we replicate our previous observations with natural scene images showing significantly bigger false alarm rates (35%) than manmade (24%; F(1,23)=20.34, p<.0001) with no significant overall effect on hit rates (56% hits for natural vs. 55% hits for manmade; F(1,23)=.056, p=.815). In the case of load effects, we see both a statistically significant overall increase in false alarms rates (F(2,46)=4.58, p=.015) and a slightly larger decrease in hit rates (F(2,46)=20.83, p<.0001) but no interactions with image type (FA p=.229; HIT p=.158).

It is clear that a demanding concurrent task such as counting backwards in threes has a major impact across conditions attenuating the ability to encode to the same extent both for manmade and natural scenes, an effect that is explored further in Experiment 5 where we revert to the door scenes studied in Experiment 3. By reverting to door scenes, we can hold constant semantic variability and directly control visual distinctiveness by manipulating level of detail in each image.

Experiment 5 & 6

Subjects and Design

Experiment 5&6 involved studying the effect of concurrent load on the door scenes used in Experiment 3. In this case, however, we were constrained by the potential size of the stimulus set (that used in Experiment 3) which restricted the number of conditions we could compare. In this experiment we circumvent this, by dividing a total of 48 new participants (30 female, aged 18-30) into two separate groups of 24, each of which was tested under two different loads in the two experiments. Experiment 5 had observers at a low load condition simply repeating out loud utterance of the three-digit numbers while in the high load condition we asked them to count back out loud in threes. Since it could be argued that simply repeating a three digit number involves a real if light cognitive load, in Experiment 6 we introduced a baseline condition without any additional load, simply instructing observers to ignore the three-digit number, in this case comparing this with the load imposed by counting back out loud in ones. For both experiments and across load conditions the observers were asked to concurrently encode the door stimuli.

Each participant was tested on two sets of 200 door scenes of which half were original doors that had idiosyncratic redundant details removed. Each study phase with two conditions of load intermixed was followed immediately by a test session comprising 200 stimuli of which half were old and half new. Again, responding was by key press and was self-paced. Half the participants began with the lower load and half with the higher. Participants had not taken part in the earlier studies, gave informed consent, had normal color vision and visual acuity and were rewarded with either credit points or a small cash payment.

Results

The results of the two experiments were analysed separately. The results of Experiment 5 involving repetition of the three-digit number and counting backwards in threes are shown in Figure 6a. An ANOVA indicated a significant main effect of concurrent load (F(1,23) = 22.67, p <.001), but no significant main effect of detail (F(1,23) =2.69, p = .115), together with a significant interaction between load and detail (F(1,23) = 6.84, p = .02). However, both the absence of a significant detail effect and the presence of an interaction probably reflect a floor effect in the count back in threes condition. Planned comparison between the two repetition conditions supports this indicating a significant effect of detail (t(46) = 4.86, p<.001), as does the absence of a reliable difference as a function of detail within the condition involving counting back in threes (t(46) = 1.85, p=.07). The two repeated measure ANOVAs in which we analysed the effect on FA 's and HIT's separately of concurrent load and level of image detail showed a significant difference between levels of image detail on FA's (F(1,23)= 9.99, p=.004) but not on HITs (F(1,23)= 1.18, p=.228; Figure 6b). Concurrent load affected both FA's (F(1,23)=4.32, p=.049) and the HITs (F(1,23)=66.28, p<.0001), but its significant interaction with the level of detail in an image was evident only in the hit rate (F(1,23)=16.97, p<.0001).

Figure 6

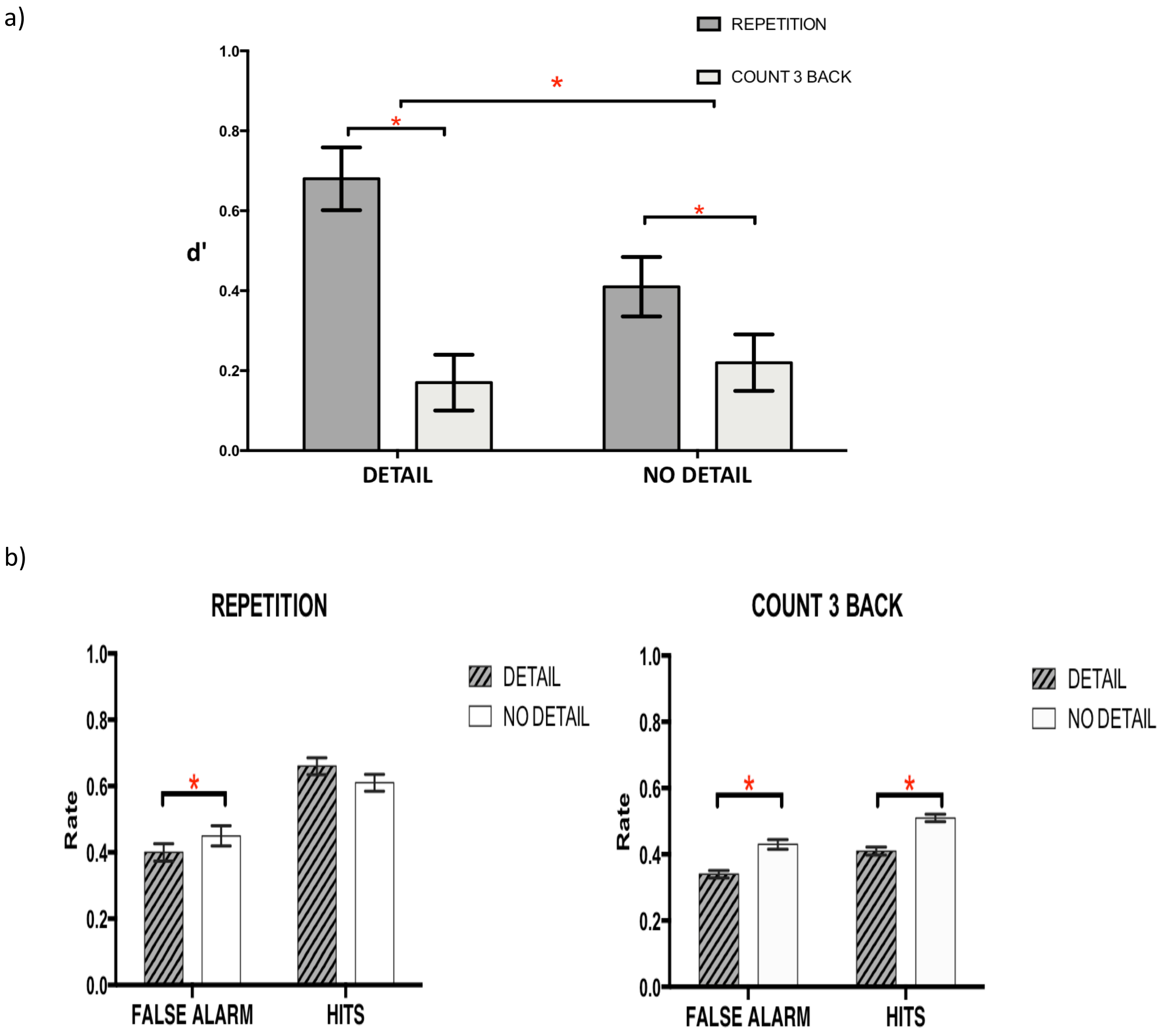


Figure 6. Performance on visual recognition memory in Experiment 5 by load and image type. a) Data from Experiment 5 across two different loads and two different image types (doors with detail vs. doors with no detail). b) Histograms of False alarm and Hit rates in Experiment 5 separated by load and image type (doors with detail vs. doors with no detail). *p<.05

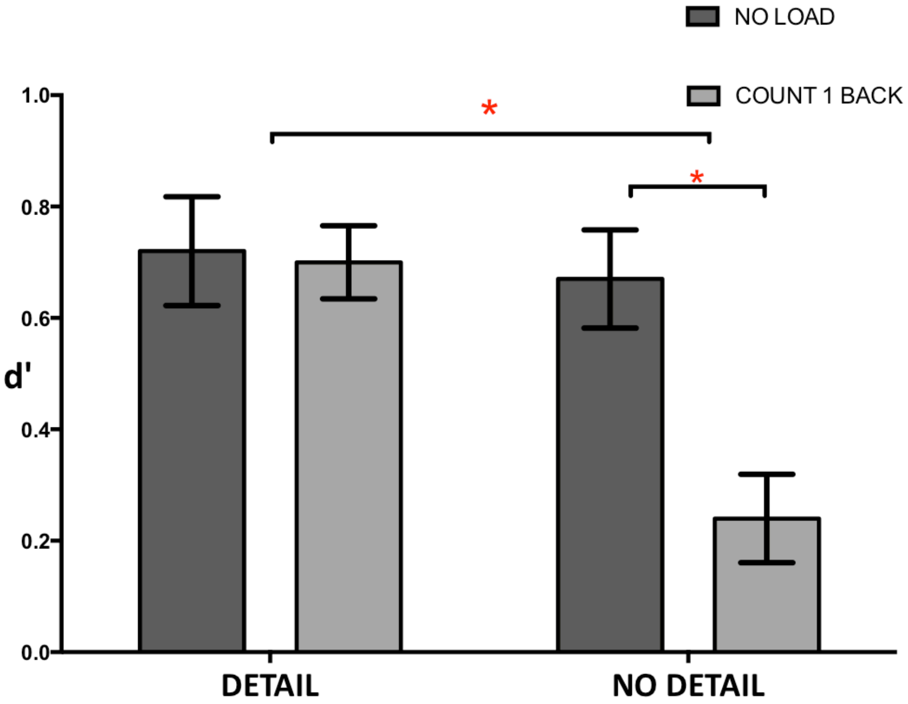
The results of Experiment 6, which involved a baseline condition of ignoring the three-digit number and of higher load of counting back in ones are shown in Figure 7a. An ANOVA indicated a significant main effect of concurrent load (F(1,23) = 8.39, p = .008), and level of detail F(1,23) = 13.27, p = .001), together with a significant level of detail by load interaction (F(1,23) =6.67, p =

.016). Multiple comparisons-corrected analysis indicated a significant effect of load for the condition with detail removed (t(23) =3.66, p = .0013) but not for the high detail condition (t(23) = .32, p = .08). A separate ANOVA on just the false alarm rates found significant effect of level of detail (F(1,23)=9.29, p=.006; see Figure 7b) that also interacted with concurrent load (F(1,23)=6.30, p=.020), where there was significantly higher rate of false alarms for images with low detail (46% without detail vs. 34% with detail) only in the count back by one concurrent load condition but not in the no load condition (39% without detail vs. 40% with detail). The hit rate was not significantly affected by level of image details (F(1,23)=.38, p=.543) but was by concurrent load (F(1,23)=15.78, p=.001) with higher load resulting in reduction in HIT's by on average 10%.

Graphs in Figure 6 a & 7a show data from both experiments 5 & 6 in which only the most difficult counting condition impairs performance on the detailed door scenes, whereas even counting back in ones impairs performance when detail is removed.

Figure 7

a)



b)

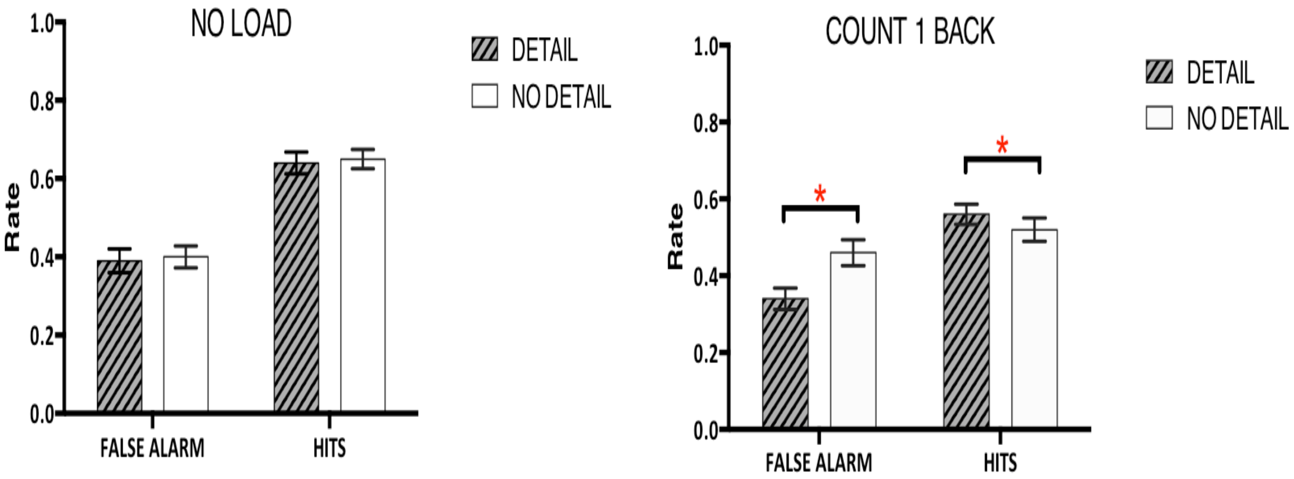


Figure 7. Performance on visual recognition memory in Experiment 6 by load and image type. a) Data from Experiment 6 across two different loads and two different image types (doors with detail vs. doors with no detail). b) Histograms of False alarm and Hit rates in Experiment 6 separated by load and image type (doors with detail vs. doors with no detail). *p<.05

Discussion

Experiments 4 to 6 are consistent in showing that the standard concurrent task of counting backwards in threes substantially impairs visual LTM for all types of material used to varying degrees, whereas only visual material that has reduced number of idiosyncratic details is sensitive to lower loads. For which even the simple requirement to repeat continuously an auditorily presented three-digit number is sufficient to produce a significant decrement in performance. Thus, we see a multifaceted interaction between influence of concurrent load and the nature of the material encoded.

The dramatic difference in performance from counting back in threes (Exp. 4 complex scenes d'=.41, Exp. 5 door scenes d'=.22) rather than in ones (Exp. 4 complex scenes d'=.76, Exp. 6 door scenes d'=.64) or repetition (Exp. 5 door scenes d'=.50) regardless of encoding material condition suggests a major switch of attention from encoding the exterior visual stimuli to the demanding executive processing task, although the fact that performance was significantly above chance in all conditions indicates that participants could remember some items even under the most demanding conditions.

Experiments 4 to 6 all show that recognition memory is substantially reduced by a demanding concurrent working memory load. This could of course simply reflect a failure to attend to the pictures because of the demands made by the concurrent task. For example, performance accuracy on the concurrent task remained high across experiments and loads; Experiment 4 performance on the concurrent low load task was 98% and high load task 91 % whereas in Experiment 5 and 6 for the concurrent two low load tasks (repetition 99% and counting back by one 97%) and high load task was 90%. The fact that load interacted marginally with type of image encoded (interaction p=.059, two-tail) in Experiment 4 is consistent with the assumption that the demanding concurrent task is still simply reducing attention rather than qualitatively changing the method of encoding. There is a clear interaction between the nature of the material and concurrent task in Experiment 6, but this appears to reflect a floor effect under the most demanding condition in Experiment 5. We also see that the load has the most consistent effect on increasing the difference in false alarms rates across different material to be encoded rather than affecting the hit rate indicating problems in differentiating between exemplars when idiosyncratic detail is impoverished. One problem in interpreting the effect of a concurrent load is to know exactly why it impairs performance. The simplest assumption is probably that as the concurrent load increases, the amount of attention left to process the images to be remembered is decreased. It is perhaps worth noting that the massive memory paradigm studies typically use relatively long exposures ranging between 3s and 10s.

General Discussion

At a theoretical level, the recent interest in long-term memory for complex scenes has been influenced by developments in computer-based models of object and scene recognition and as such have principally focused on the nature of the scene rather than the role of the rememberer (Brady et al, 2008; 2011). The current exploratory study aims to complement this approach by focusing on two broad attentional factors that may potentially influence visual LTM, namely strategy and available attentional capacity, both of which have been extensively investigated in the case of verbal LTM (Baddeley, Lewis, Eldridge & Thomson, 1984, Craik & Lockhart 1972; Murdock 1965). We do so across a range of visual stimuli, from complex manmade scenes to door scenes in which distinctive features have been removed. This resulted in broad answers to the two general questions together with unexpected features that have promising implications for further developing theories of visual LTM. We will discuss our two major questions in turn, in each case focusing initially on the effect of the relevant variables on overall performance as measured by d', before looking separately at analyses based on hit rates and false alarms.

Table 1

Intention to Learn (η_p^2)

	Intention			Material		
	d'	Hits	FAs	d'	Hits	FAs
Exp. 1	.03	.02	.00	.80***	.36**	.72***
Exp. 2	.01	.01	.00	.45***	.06	.64***
Exp. 3	.10*	.01	.01	.66**	.05	.52**

Table 1. Effect sizes, expressed as partial eta squared, of intent to learn (strategy) and type of material (visual complexity) on memory performance (d'), Hit and False alarm rates.

*p<.05,**p<.01,***p<.001 Statistically significant effect.

Data from Experiments 1, 2 and 3, concerned with intention to remember, are summarised based on effect sizes [η_p^2] in Table 1, with the left-hand side of the table concerned with the effect of intention to learn, and the right hand with that of difference in material between the stimuli with greater, and those with less visual and/or sematic complexity. We express this later difference in terms of complexity, an issue that we will discuss later. The first point to note is that when measured by d', there is no effect of the instruction that memory of the stimuli will be tested later in either Experiment 1 or 2, although importantly, in the case of door scenes (Exp.3) there is a weak effect primarily carried and qualified by an interaction whereby the most difficult stimuli, doors with idiosyncratic features removed, show an effect of intention to remember. Columns 2 and 3 refer to analyses based on hit rates and false alarms respectively. Notably, while only one of the three experiments shows a material-based effect for hit rates, all three show a substantial effect based on false alarms. It is notable that the one condition showing an effect on hit rate also shows an effect on false alarms that is double in its magnitude. This condition shows the highest level of performance with a d' prime score on the manmade scenes of 2.54, in contrast to a d' prime of 0.79 for the most difficult doors condition. In contrast, all three studies show clear effects on false alarm rate, with the effect in Experiment 1 being double that of hit rate effect.

The overall pattern of results therefore suggests that intention to learn has remarkably little effect on overall performance, except for the most difficult stimuli, door scenes with prominent features removed. In the case of Experiment 1, it could be argued that the task of judging pleasantness involves the relatively deep level of processing that incidentally results in the encoding of the relevant features for later recognition (Baddeley & Hitch, 2017). However, when the encoding task is switched in Experiment 2 from "deep" processing to a relatively "shallow" search for an extraneous stimulus, although performance declines substantially, the pattern is essentially the same, suggesting the need to search actively for discriminative features, as potentially prompted by intention to remember becomes necessary only under conditions when more obvious discriminative features are removed.

In the interests of studying the generality of any findings across level of difficulty, we included in all our studies material from two subcategories that were expected to differ in level of difficulty. As expected, this influenced performance, but unexpectedly the effect was principally on false alarms rather than hit rate. Experiments 4 to 6 also include these contrasting sets of material and hence offer a potential replication of our unexpected findings.

Table 2

Effect of Load (η_p^2)

	Load Effect			Material		
	d’	Hits	FAs	d’	Hits	FAs
Exp. 4	.43***	.48***	.17*	.43*	.00	.47***
Exp. 5	.50***	.74***	.16*	.11	.05	.30**
Exp. 6	.26**	.41**	.00	.36**	.02	.29**

Table 2. Effect sizes, expressed as partial eta squared, of cognitive demand (executive attention) and type of material (visual complexity) on memory performance (d’), Hit and False alarm rates.

*p<.05, **p<.01, ***p<.001 Statistically significant effect

Having found that strategic factors do not appear to have a strong impact on visual LTM, our second question asked whether this was because encoding was automatic without the need for executive capacity. We addressed this using a dual task to absorb executive attention, a process that has proved fruitful in studies of verbal working memory (Baddeley & Hitch, 1974; Baddeley, Hitch & Allen, 2009). The results of Experiments 4 to 6 are summarised in Table 2 where again, the effects of the principal variable, concurrent load, is shown on the left most columns and that of material on the right, with results again represented in terms of effect sizes (η_p^2). Considering first the d' prime measures, in contrast to Experiments 1 to 3, concurrent load clearly has a highly consistent effect on performance with higher loads leading to poorer subsequent retention. When these results are broken down into hits and false alarm rates, again in contrast to Experiments 1 to 3, we observe clear effects of load on hit rate, the likelihood of detecting a previously presented stimulus, whereas a small effect on false alarms is found in Experiment 4 and 5 but not in 6. Experiment 5, based on door scenes, shows a small effect on false alarm rate, although it should be noted that performance in Experiment 5 was virtually at chance in the count back by three conditions. Hence, while this study gives some idea of the fragility of visual memory under distraction, its detail should be treated with caution. What can be concluded from Experiments 4 to 6 however, is that visual memory is clearly sensitive to distraction from concurrent verbal tasks, with more demanding tasks creating greater impairment. This contrasts with the lack of impact of intention to remember in Experiments 1 to 3.

Our six experiments were exploratory in nature and our results did not attempt to test specific hypotheses concerning the nature of memory for scenes. We began with a puzzling observation namely that while on the one hand people appear to be able to remember up to 10,000 images with a recognition rate of over 80%, on the other hand, lists of only 12 pictures of doors lead to error rates that are sufficient to reliably detect the effects of age, schizophrenia and relatively mild degrees of brain damage (Baddeley Emslie & Nimmo-Smith, 1994). While theorising about memory extends back for over a century, we still do not have a good explanation of this apparent paradox. This partly reflects the comparative neglect of investigation of visual long-term memory until relatively recently, largely involving investigators with an interest in vision and the development of computational approaches to object recognition. Theories of object recognition have been proposed as offering a possible approach to understanding visual LTM but run into a significant problem. The purpose of object recognition theories is to account for the way in which a wide range of objects or scenes such as cups or forests potentially involving variations in shape, size and colour can each be mapped on to a specific concept. The memory problem is the reverse of this, that of remembering a specific example of a cup or a forest scene and distinguishing it after a delay from an apparently similar cup or scene. The fact that it can be done successfully after a delay indicates some form of memory trace, though not of course necessarily the retention of the whole scene. This raises the question of how a theory based on moving from the specific to the general as in object recognition can be adapted to account for successful retention of the specific object or scene. Classic theories of LTM also imply some form of storage of information but have typically been developed using verbal material and interpreted in terms of unspecified "features" (Estes, 1997; Nairne, 2002).

We would however like to offer a possible explanation of our broad pattern of results by combining the object-based and verbal approaches as follows. While the end point of an object recognition programme may be a single node, this is arrived at via a number of earlier patterns of activation which may then be discarded. We suggest however that if these patterns are primed in some way, so as to leave a potential trace, then they may form the basis of the features assumed by the verbal memory models. We use the term "priming" in its broadest sense as a phenomenon that occurs widely throughout the nervous system whereby the operation of a process influences its subsequent use, leaving its precise nature open. Memory implies that something persists within the nervous system, with priming probably reflecting the most basic form of such persistence. We suggest that presentation of a test stimulus will re-activate a pattern of primed features, those involved in object recognition, and that this primed array will correspond to a greater or lesser degree to the original and that this gives rise to a greater or lesser sensation of familiarity.

The problem then is to successfully evaluate this familiarity level. The difficulty of this will of course depend on the nature of the material to be remembered. Some test items may be clearly different at a broad semantic level, an effect reflected in the fact that performance improves as more semantic categories are included in a list of visual scenes, resulting in a wider range of available features (Konkle et al 2010a). However, as we ourselves show, performance can be relatively high even with a single rather narrow category such as doors, suggesting the need to use more detailed levels in order to access familiarity. Furthermore, as Nairne (2002) points out, the crucial factor is not the sheer number of features but their "diagnosticity", their capacity to rule out items that were not previously presented. Hence, although a scene may be stored on the basis of many features, if most of these are shared with the "new" items in the set, they will not be helpful, whereas a single non-matching feature might prove crucial. In this connection it is notable that memory for lists of simple objects on a blank background is relatively high, despite the fact that the number of features present are likely to be considerably less than those for natural scenes (Konkle et al 2010b), suggesting that it may be the proportion of diagnostic features rather than number of features encoded.

We use this broad framework to account for our three major findings. These comprise a lack of influence from intention to remember, the impact of attention-demanding concurrent tasks and thirdly the consistent tendency for our more difficult material to lead to an increase in false alarms rather than a decrease in correct detections. We suggest first of all that the relative unimportance of intention to remember, like the absence of "deep" processing effects in the Baddeley and Hitch studies reflects the fact that visual stimuli potentially comprise a rich array of visual features that are encoded rapidly and in parallel. It is known from studies using rapid serial visual presentation (RSVP) that participants are able to extract semantic information extremely rapidly, selecting for example a beach scene from a range of rapidly presented scenes of a different kind (Potter, Wyble, Hagmann, & McCourt, 2014). Likewise, a 20-ms masked exposure of a single image is enough to categorize the basic (e.g., lake vs. forest) or superordinate (e.g., natural vs. urban) level of a scene with above-chance accuracy (Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). What is more, observers are capable of rapidly extracting information about multiple categories, even if they do not know the target category in advance (Evans, Horowitz, & Wolfe, 2011). It seems likely therefore that the visual system is capable of processing and potentially storing very large amounts of information extremely rapidly. The task of recognition involves deciding whether a particular stimulus has or has not been seen before. This does not of course mean in the cases of visual stimuli that every feature of a whole complex picture has been encoded and stored, although successful recognition does indicate that sufficient discriminative features have been retained to allow a judgment based on either familiarity or potentially on recollection of the experience of encoding that item.

It is only when stimuli are impoverished, and the potentially diagnostic features are limited, that strategy becomes important, presumably by encouraging the rememberer to focus more effectively on features that might potentially be diagnostic. It might be tempting to assume that the rapid and parallel processing of the relevant features in a scene is entirely automatic. However, while this could be the case, the clear impact of our attentionally demanding concurrent tasks on retention suggests that encoding into LTM is not automatic but depends on some form of intermediate processing at an attentionally limited stage. Within the multicomponent working memory model, this would be likely to involve the interaction between the visuospatial sketch pad, the episodic buffer and the central executive (Baddeley, 2012; Baddeley, Allen & Hitch ,2011).

Further constraints on possible theories and our framework are provided by the previously discussed pattern of false alarm rates with a tendency for differences in difficulty across materials to be reflected in increased false alarm rates rather than reduced probabilities of detection. Given its unexpected nature, this observation should clearly be treated with caution, but the fact that it is repeated across the two sets of experiments involving a range of different paradigms, different procedures, different materials and different levels of performance would seem to demand an explanation. The standard explanation of false alarm rates is in terms of criterion effects with a lax criterion leading to more false alarms (Swets, Tanner & Birdsall, 1961). Given that the two levels of difficulty within each study are varied randomly within the sequence of items presented, this would imply a switch in criterion from one item to the next (Singer, 2009). Why should this be?

We argue that this pattern suggests that the process of recognition may involve two levels, one potentially based on familiarity and a second checking stage that appears to reflect a need to use further diagnostic detail to decide in cases of doubt. Schacter, Israel and Racine (1999) report a similar error pattern to our own, of retention difficulty reflected in false alarm rate rather than correct detections, although in their case, based on a verbal recognition study. They suggest that this checking process which they refer to as a “distinctiveness heuristic” may be based on episodic memory, the capacity to recollect the encoding experience (Tulving, 1985). The distinction in recognition memory between simply “knowing” that an item has been included in the memory set and “remembering” the experience has proved to be a robust one in studies of both healthy participants (Gardiner & Java,1991) and patients (Brandt, Gardiner, Vargha-Khadem, Baddeley & Mishkin,2009). This interpretation depends on what is encoded, and our data suggest that this in turn maybe influenced by instruction to remember and by available attentional capacity. The interpretation by Schacter et al. suggest further involvement at the retrieval stage. It does of course remain to be seen how robust and general the observed error pattern proves to be, and to what extent it fits the interpretation proposed by Schacter et al (1999).

The two-level processing, we propose, sees the first stage based on gist, followed by a later more detailed process based on individual features. The gist level may allow a rapid relatively automatic judgement, with the more detailed stage used as a further source of evidence when in doubt. We suggest that this stage may tend to accept a riskier or rather a more lenient criterion to be adopted. The gist level may involve not only the rapid visual encoding of the stimuli but also semantic features such as category membership; as Konkle et al. (2010a) demonstrated, the more categories of visual stimulus comprises, the higher the level of performance. At this level one encodes very general features (e.g. the scene is an indoor scene of a kitchen rather than a kitchen scene with a kitchen island, 5 burner stove and green refrigerator) which are easily confusable with the features of the lures of the same semantic category. All of our experiments involve stimuli selected from two categories, categories that are more clearly defined in the case of scenes than for door stimuli, potentially resulting in the higher overall level of performance observed.

Finally, it is important to note that our explanation is currently limited to recognition memory. Data from studies using recall can appear to be very different. A recent study by Baddeley, Atkinson, Kemp & Allen (in press) was based on just four distinctive door scenes (domestic, gate, factory and church) each comprising five clearly discriminable features tested by cued feature-based cued recall, (e.g. What colour was the wall surrounding the church door?). Brief presentation, up to 5s per scene led to poor performance and a full 10s presentation per scene with an immediate check that the features had been successfully encoded proved necessary before a reasonable level of long-term retention was achieved. This is not of course to suggest that totally different memory systems are involved in visual recognition and recall, but suggests that relying on a limited range of paradigms may lead to conclusions that are premature and potentially misleading, as in the case of the suggested massive capacity of visual long-term memory perhaps?

In conclusion, our results indicate that there might be two levels of processing of complex visual material into long-term memory. The first relatively rapid and unaffected by intention to remember results in the, processing of the multiple features available in a complex visual scene (e.g. the gist of the scene). This is sufficient for successful recognition in many cases and consequentially in encoding the gist of that stimulus. Where this is not sufficient, a second recognition process is based on retention of more detailed, distinctive and diagnostic features of the visual stimulus. This level is sensitive to the strategy adopted and to the availability of attentional resources. It allows for discrimination between exemplars from the same category resulting in a lower susceptibility to false positives. The multi-level nature of encoding visual material into long term memory comes to light only when visual features become impoverished. The second encoding process involves a less strict decision criterion with difficulty level shown in increased false alarm rates rather than reduced detections. Both levels however demand a degree of attention and cannot tolerate disruption from demanding concurrent cognitive activity.

Acknowledgment

We are grateful to Drs. Vogt and Magnussen for permission to use their stimuli.

References

Baddeley, A. (2012). Working memory, theories models and controversy. *The Annual Review of Psychology, 63*, 12.1–12.29.

Baddeley, A.D., Allen, R.J., & Hitch, G.J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49,* 1393-1400.

Baddeley, A. D., Atkinson, A. Kemp, S. & Allen, R. (in Press). The problem of detecting long-term forgetting: Evidence from the Crimes Test and The Four Doors Test. Cortex.

Baddeley, A.D., Emslie, H. & Nimmo-Smith, I. (1994). Doors and People: a Test of Visual and Verbal Recall and Recognition. Bury St Edmunds, Suffolk, Thames Valley Test Company.

Baddeley, A. D., Eysenck, M., & Anderson, M. C. (2014). *Memory* (2nd ed.). Hove: Psychology Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (2017). Is the Levels of Processing effect language-limited? *Journal of Memory and Language*, *92*, 1-13.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory & Language, 61*, 438-456.

Baddeley, A. D., Hitch, G. J., Quinlan, P. T., Bowes, L., & Stone, R. (2016). Doors for memory: A searchable database. The Quarterly Journal of Experimental Psychology, *69*, 1-18.

Baddeley, A. D., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113*, 518-540.

Baddeley, A. D., Scott, D., Drynan, R., & Smith, J. C. (1969). Short-term memory and the limited capacity hypothesis. *British Journal of Psychology, 60*, 51-55.

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*, 1-4.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of National Academy of Sciences U S A, 105*, 14325-14329.

Brandt, K. R., Gardiner,J.M., Vargha-Khadem, F., Baddeley, A. D. & Mishkin,M. (2009).Impairmant of recollection but not familiarity in a case of developmental amnesia. Neurocase, 15(1), 60-65.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Castelhano, M., & Henderson, J. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017-1040.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing. A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*, 671-684.

De Renzi, E., & Nichelli, P. (1975). Verbal and non-verbal short-term memory impairment following hemispheric damage. *Cortex, 11*, 341-353.

Estes W.K. (1950) Toward a statistical theory of learning. Psychological Review, 57, 94-107.

Evans, K.K., Cohen, M., Tambouret, R., Horowitz, T., Kreindel, E. & Wolfe, J.M. (2010) Does Visual Expertise Improve Visual Recognition Memory? Attention, Perception & Psychophysics, 73 (1), 30-35.

Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). When Categories Collide Accumulation of Information About Multiple Categories in Rapid Scene Perception. *Psychological science*, *22*(6), 739-746.

Fernandes, M., & Guild, E. (2009). Process-specific interference effects during recognition of spatial patterns and words. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *63*(1), 24.

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory & Cognition*, *19*(6), 617-623.

Greene, M. R., & Oliva, A. (2009). The briefest of glances The time course of natural scene understanding. *Psychological Science*, *20*(4), 464-472.

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*, 1469-1482.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision research*, *47*(26), 3286-3297.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Scene memory is more detailed than you think: The role of catogories in visual long-term memory. *Psychological Science, 21*, 1551-1556.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General, 139*, 558-578.

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 1-11.

Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* (Vol. 1, pp. 328-372). New York: Academic Press.

Murdock Jr, B. B. (1960). The distinctiveness of stimuli. *Psychological review*, *67*(1), 16.

Murdock, B. B. (1965). Effects of a subsidiary task on short-term memory. *British Journal of Psychology*, *56*(4), 413-419.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*, 389-395. doi:10.1080/09658210244000216

Öhman, A. (2009). Of snakes and faces: An evolutionary perspective on the psychology of fear. *Scandinavian Journal of Psychology, 50*, 543-552. doi:10.1111/j.14679450.2009.00784.x

Paivio, A. (1971). *Imagery and Verbal Processes*. London: Holt Rinehart and Winston.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, *76*(2), 270-279.

Rey, A. (1964). *L'Examen Clinique en Psychologie*. Paris: Presses Universitaires de France.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and language*, *40*(1), 1-24.

Singer, M. (2009). Strength-based criterion shifts in recognition memory. *Memory & Cognition*, *37*(7), 976-984.

Snodgrass, J. G., & Corwin, J. (1988) Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology General, 117*, 34-50.

Standing, L. (1973). Learning 10000 pictures. *Quarterly Journal of Experimental Psychology, 25*, 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73-74.

Swets, J. A., Tanner Jr, W. P., & Birdsall, T. G. (1961). Decision processes in

perception. *Psychological review* , *68*(5), 301.

Tulving, E. (1985). How many memory systems are there? The American Psychologist. 40, 385 398.

Vallar, G., & Shallice, T. (Eds.). (1990). *Neuropsychological impairments of short-term memory*. Cambridge: Cambridge University Press.

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology, 54*, 298-303. doi: 10.1027/1618-3169.54.4.298

Warrington, E. K. (1984). Recognition Memory Test. Windsor UK: NFER-Nelson.