Cognition 180 (2018) 24–37

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Intention, attention and long-term memory for visual scenes: It all depends on the scenes

Karla K. Evans*, Alan Baddeley

Department of Psychology, The University of York, York, North Yorkshire, UK

ARTICLE INFO

Keywords:
Visual episodic memory
Executive attention
Intent to learn
Encoding
Working memory

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 4–6 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.

1. 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 (De Renzi & Nichelli, 1975; Vallar & Shallice, 1990).

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, 1973; Standing, Conezio, & Haber, 1970). 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

* Corresponding author.
E-mail addresses: karla.evans@york.ac.uk (K.K. Evans), a.baddeley@york.ac.uk (A. Baddeley).

https://doi.org/10.1016/j.cognition.2018.06.022
Received 1 February 2018; Received in revised form 26 June 2018; Accepted 27 June 2018
0010-0277/ © 2018 Elsevier B.V. All rights reserved.

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 to 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 (Baddeley & Hitch, 2017; Castelhano & Henderson, 2005).

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 (Baddeley, Scott, Drynan, & Smith, 1969; Murdock, 1960). 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.

2. Intent to learn and encoding

One way to probe the effects of encoding method is to test if it requires intent. Thus, in Experiments 1–3 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 et al., 2010) suggests broadly similar d′ scores across studies using yes/no, two-alternative and four-alternative forced choice.

Download English Version:

https://daneshyari.com/en/article/7285026

Download Persian Version:

https://daneshyari.com/article/7285026

Daneshyari.com