The neural basis of precise visual short-term memory for complex recognisable objects

Michele Veldsman [a,*], Daniel J. Mitchell [b], Rhodri Cusack [c,d]

[a] Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[b] Medical Research Council Cognition and Brain Science Unit, Cambridge, UK
[c] Brain and Mind Institute, University of Western Ontario, London, Canada
[d] Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

ABSTRACT

Recent evidence suggests that visual short-term memory (VSTM) capacity estimated using simple objects, such as colours and oriented bars, may not generalise well to more naturalistic stimuli. More visual detail can be stored in VSTM when complex, recognisable objects are maintained compared to simple objects. It is not yet known if it is recognisability that enhances memory precision, nor whether maintenance of recognisable objects is achieved with the same network of brain regions supporting maintenance of simple objects.

We used a novel stimulus generation method to parametrically warp photographic images along a continuum, allowing separate estimation of the precision of memory representations and the number of items retained. The stimulus generation method was also designed to create unrecognisable, though perceptually matched, stimuli, to investigate the impact of recognisability on VSTM. We adapted the widely-used change detection and continuous report paradigms for use with complex, photographic images.

Across three functional magnetic resonance imaging (fMRI) experiments, we demonstrated greater precision for recognisable objects in VSTM compared to unrecognisable objects. This clear behavioural advantage was not the result of recruitment of additional brain regions, or of stronger mean activity within the core network. Representational similarity analysis revealed greater variability across item repetitions in the representations of recognisable, compared to unrecognisable complex objects. We therefore propose that a richer range of neural representations support VSTM for complex recognisable objects.

1. Introduction

Visual short-term memory (VSTM) enables us to operate in our rich and dynamic visual environment. In the last decade, there have been many advances in theoretical understanding of VSTM (Luck and Vogel, 2013) but these have largely been based upon experiments using simple abstract stimuli such as coloured patches and geometric shapes that lack important facets of complex visual objects in the environment (Brady et al., 2011). Most saliently, objects in the environment are frequently recognisable, allowing rich information to be retrieved from long-term memory, such as their category, prototypical form, and semantics. But, does this influence our memory for their visual features?

Recently, Brady et al. (2016) demonstrated quantitatively different VSTM capacity for complex real world objects compared to simple colours. More complex items could be stored with finer detail than simple colours when encoding time was increased. But simple colours and complex objects differ in both semantic and perceptual complexity, therefore careful control of visual features is important in assessing the quality and capacity of VSTM for complex objects and the impact of recognisability. Similarly, it remains to be determined how complex, recognisable objects in VSTM are represented in the brain compared to perceptually matched, complex, but unrecognisable items.

Brady et al. (2016) showed that maintenance of complex items was associated with higher amplitude contralateral delay activity (CDA), an electrophysiological marker that indexes the amount of visual information actively stored. Contrary to the expectation that episodic long-term memory would support the maintenance of complex recognisable objects, Brady et al. (2016) argued that the same mechanism supporting maintenance of simple objects, indexed by the CDA, also supported maintenance of complex objects. This raises the question as to whether

* Corresponding author. Nuffield Department of Clinical Neurosciences, University of Oxford, Level 6, West Wing, John Radcliffe Hospital, UK.
E-mail address: michele.veldsman@ndcn.ox.ac.uk (M. Veldsman).

the same spatial network of regions recruited in the maintenance of simple objects is also sufficient to support maintenance of complex recognisable objects when they are perceptually matched. A meta-analysis of change detection tasks showed activity across a network of frontal and parietal regions in the encoding and maintenance of simple objects (Linke et al., 2011). VSTM for complex recognisable objects, with their semantic associations, may recruit the same regions to a greater extent or recruit additional regions, such as those associated with long term memory. Alternatively, recognisability may modulate fine-scale activity patterns within the core network.

Our first goal was to determine how VSTM for complex objects is affected by how recognisable they are. To study this, we created sets of variably recognisable stimuli that were matched in their visual properties. We used a warping procedure that made photographs of objects more difficult to recognize, but across the set gave indistinguishable distributions of responses in computational models of early visual processing (Stojanoski and Cusack, 2014).

Our second goal was to understand the neural mechanism through which recognisability affects VSTM. We hypothesised that visual regions may be tuned for recognisable objects, which may allow for more efficient neural codes that are more effectively maintained in VSTM. A second hypothesis was that the recognition of an object could recruit additional regions in the hierarchically organised ventral visual stream (DiCarlo et al., 2012). These additional regions might directly contribute to VSTM, or might indirectly support representations in earlier visual regions.

To preview the results, across three experiments (total N = 88) we demonstrate that the visual features of recognisable objects are remembered with higher precision than those of unrecognisable objects. However, we found no consistent evidence that this clear behavioural benefit was supported by increased regional brain activation, or recruitment of additional brain regions. Instead, multivariate pattern analysis revealed greater variability in the representation of recognisable objects, suggesting a wider representational space may support VSTM for complex, recognisable objects.

2. Experiment 1

In experiment 1 we measured the effect of recognisability on VSTM, and the neural basis of this effect. We first compared the detection of large (cross-category) versus small (within-category) changes. This manipulation has previously been applied a number of times to assess the precision of a VSTM representation from the likelihood of an item being remembered at all (Awh et al., 2007; Scolari et al., 2008; Barton et al., 2009).

2.1. Methods

2.1.1. Participants

Eighteen participants (9 male, aged 18–46, mean age 28) gave informed consent, approved by the Cambridge Research Ethics Committee, and were paid for taking part. Participants had no history of psychological or neurological health problems and reported normal or corrected to normal vision.

2.1.2. Stimuli

Forty colour photographic images of real world objects (21 depicting inanimate objects) were selected from the stimuli used in Kriegeskorte et al. (2008b). Stimuli were then manipulated in Matlab (MathWorks, 2009B). A diffeomorphic transformation was applied to each photographic image to parametrically degrade its recognisability. Diffeomorphic warping was preferred over procedures such as phase, box or texture scrambling that are also designed to remove recognition, because it provides better matching of visual properties (Stojanoski and Cusack, 2014). Like bending a rubber sheet, the diffeomorphic transformation maintains a 1:1 mapping between each point in the source and a point in

the target, in a continuous transformation across space, without replacement or duplication. The transformation is also smooth and reversible (see Stojanoski and Cusack (2014) for full details of the transformation).

2.1.3. Change detection paradigm

We used the common change detection paradigm to measure VSTM (Cowan, 2001; Scolari et al., 2008). A single sample item was presented centrally for 0.25s spanning a visual angle of 3.70° on a uniform grey background (Fig. 1a). The sample was either intact (recognisable) or heavily distorted (unrecognisable). Memory load was fixed to a single item. The sample was followed by a jittered maintenance period (selected from a random distribution between 1 and 9 s) in which a uniform grey background was displayed. A probe was then presented centrally for 1.5s, which matched the sample item on half of the trials and changed on half of the trials. Participants made a 'same' or 'different' response with a response button box in the scanner. Response mapping was counterbalanced across participants. A blank inter-trial-interval followed the probe, whose duration was selected from a random distribution between 1 and 9 s. The jittered inter-trial and inter-stimulus periods were designed to allow separation of the encoding, maintenance and response periods (Rowe and Passingham, 2001).

To distinguish between complete forgetting, imprecise remembering, and precise remembering, we manipulated the type of change to be detected (Scolari et al., 2008). In cross-category (CC) trials the identity of the stimulus changed between the sample and probe, with the level of recognisability (degree of diffeomorphic distortion) maintained. In within-category (WC) trials the level of recognisability changed between the sample and the probe, with the identity of the sample maintained. 40 CC and 40 WC trials were blocked and the order of blocks was counterbalanced across participants. Within each block, change/no-change trials were counterbalanced with the recognisability manipulation, and presented in a random order. VSTM capacity was estimated using Cowan's formula (*hits-false alarms)/set size*; Rouder et al., 2011) which gives a measure of the items-worth of information retained (K).

The task was presented to participants with a PC running the Psychophysics Toolbox Version 3 extension of Matlab (Mathworks, Natick, MA). This was projected into the scanner using a Christie (Cypress, CA) video projector at a 60 Hz refresh rate and viewed in a mirror approximately 90 mm from the participants' eyes.

2.1.4. MRI acquisition

Scanning took place at the Medical Research Council Cognition and Brain Sciences Unit, Cambridge on a 3T Siemens TIM Trio (Erlangen, Germany). For fMRI, T2*-weighted echo-planar images (EPI) were acquired (Repetition time (TR) 2s; echo time (TE) 30s; flip angle 78°; 32, 3.5 mm slices with 10% gap, 64 × 64 acquisition matrix and 3 × 3 × 3.75 mm voxel size). The first 10 s of scans were discarded to allow for T1 equilibrium. For anatomical localisation, a high-resolution T1-weighted 3D MPRAGE was acquired with TR of 2.25 s; TE of 2.99 ms; TI of 900 ms, 9° flip angle, 256 × 240 × 192 matrix size and 1 mm isotropic voxels.

2.1.5. Image processing and analysis

Automated processing software (aa, www.github.com/rhodricusack/automaticanalysis) was used to preprocess and analyse the functional imaging data (Cusack et al., 2014) in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm/). The preprocessing pipeline comprised motion correction, slice-time correction, coregistration to the individual's structural image, normalisation to MNI template space and smoothing with a 10 mm FWHM Gaussian kernel. A high pass filter with a cut-off of 128 s was applied to remove low frequency noise. Regressors of interest modelled recognisable and unrecognisable target trials and CC and WC change trials in the encoding, maintenance and response periods of the task. Events were convolved with the canonical hemodynamic response function and a