Cognition 159 (2017) 85–96

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Hierarchical organization in visual working memory: From global ensemble to individual object structure

Qi-Yang Nie [a,*], Hermann J. Müller [a,b], Markus Conci [a]

[a] Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany
[b] Department of Psychological Sciences, Birkbeck College, University of London, London, United Kingdom

ARTICLE INFO

Article history:
Received 21 June 2016
Revised 14 November 2016
Accepted 23 November 2016
Available online 1 December 2016

Keywords:
Change detection
Visual working memory
Global precedence
Object structure
Ensemble representation

ABSTRACT

When remembering a natural scene, both detailed information about specific objects and summary representations such as the gist of a scene are encoded. However, formal models of change detection that are used to estimate working memory capacity, typically assume observers simply encode and maintain memory representations that are treated independently from one another without considering the (hierarchical) object or scene structure. To overcome this limitation, we present a hierarchical variant of the change detection task that attempts to formalize the role of object structure, thus, allowing for richer, more graded memory representations. We demonstrate that detection of a global-object change precedes local-object changes of hierarchical shapes to a large extent. Moreover, when systematically varying object repetitions between individual items at a global or a local level, memory performance declines mainly for repeated global objects, but not for repeated local objects, which suggests that ensemble (i.e., summary) representations are likewise biased toward a global level. In addition, this global memory precedence effect is shown to be independent from encoding durations, and mostly cannot be attributed to differences in saliency or shape discriminability at global/local object levels. This pattern of results is suggestive of a global/local difference occurring primarily during memory maintenance. Altogether, these findings challenge visual-working-memory (vWM) models that propose that a fixed number of objects can be remembered regardless of the individual object structure. Instead, our results support a hierarchical model that emphasizes the role for structured representations among objects in vWM.

1. Introduction

Visual working memory (vWM) enables cognitive functions to operate independently of direct retinal stimulation, with current contents in vWM supporting goal-directed behavior. However, in order to maintain a stable representation of the world, only a limited amount of sensory information of an individual's total visual input can be represented in vWM (Luck & Vogel, 2013, for a review). Hence, a major focus of studies on vWM is to describe the organizational principles by which this limited cognitive space can be used efficiently for the internal representation of visual input.

Much of the work in this regard has followed from Luck and Vogel's (1997) seminal study. They devised a change detection task in which a memory array of colored squares (varying across trials from a single square to up to 12 squares) was presented for a few hundred milliseconds (ms). Subsequent to a brief blank delay of about 1 s, a probe array was presented that contained the same items as the memory array – except (on half of the trials) for one object that was displayed in a different color. Observers were required to detect the change by giving a yes/no (two-alternative) forced-choice response. The results from these experiments indicated that participants had a vWM capacity around three to four items (Luck & Vogel, 1997). Moreover, they found that an individual's capacity did not change with the number of features that combined to form a given object. For instance, detecting a feature change was equivalent when comparing objects determined by conjunctions of four features (and where all of these features could potentially change) with objects defined by a single feature only. This observation led Luck and colleagues to propose that the capacity of vWM is (relatively) fixed: there are only a limited number of available slots, each one capable of storing a single object representation regardless of its complexity (Vogel, Woodman, & Luck, 2001).

* Corresponding author at: Allgemeine und Experimentelle Psychologie, Department Psychologie, Ludwig-Maximilians Universität, Leopoldstr. 13, D-80802 München, Germany.

E-mail address: qiyang.nie@psy.lmu.de (Q.-Y. Nie).

http://dx.doi.org/10.1016/j.cognition.2016.11.009
0010-0277/

The slot model has been challenged from at least two perspectives, and the nature of vWM capacity limits remains a topic of vigorous debate. One open question concerns the influence of object complexity. For example, contrary to the findings of Vogel et al. (2001), others have shown that vWM capacity declines with increasing object complexity (e.g., Alvarez & Cavanagh, 2004). A second challenge arises from the notion that, rather than there being a limited number of available slots, vWM capacity may depend on a single information-limited cognitive resource. Evidence for this alternative view was originally provided by Bays and Husain (2008) using a variation of the change detection paradigm. On each trial, a sample array of colored squares was presented, followed by a brief delay and a subsequent test probe. The task was to report whether the test probe was displaced to the left or the right of the corresponding item in the memory array. The results showed that performance remained near-perfect for sufficiently large displacements even when presenting rather large set sizes (e.g., a set size of 8 objects, Bays & Husain, 2008). Moreover, mnemonic precision declined monotonically as a function of memory load – an outcome expected if vWM were supported by a limited-capacity resource that requires to be distributed across more objects as the memory load increases. In this view, retaining a small number of items can be accomplished with relatively high precision; but an increase in the number of to-be-remembered items leads to a decline in the precision with which items can be remembered. This trade-off between quality and quantity of mnemonic representations implies that memory resources can be allocated flexibly among several items stored in vWM to maximize mnemonic precision, given the available resources.

Overall, these findings imply that information limits in vWM are determined both by the number and the precision of mnemonic representations (e.g., Luck & Vogel, 1997; Wilken & Ma, 2004). To accommodate this, slot models have been modified to allow for variable representational precision within a slot (Luck & Vogel, 2013; Zhang & Luck, 2008). However, one contentious question that remains is how best to explain capacity limitations: Is the amount of visual information an individual can retain in vWM limited because of a limited number of slots (i.e., caused by an absolute ceiling in performance) or because, at some point, resources have been distributed so widely that the mnemonic fidelity for any given item becomes too poor for the item to be retrievable? Moreover, vWM models as described above tend to focus on how observers encode independent features or objects from rather simple arrays of segmented geometric shapes without considering the rather complex relational structure that is usually present in the natural ambient environment.

Contrary to the simple stimulus arrays used in most vWM studies, memory for real-world scenes has been shown to depend largely on organizational principles, that is, mechanisms that impose structure on visual input. For example, when trying to remember natural scenes, the gist of that scene (e.g., a statistical summary, or ensemble representation) is encoded, in addition to the detailed information about relatively few specific objects (Conci & Müller, 2014; Hollingworth & Henderson, 2003; Oliva, 2005). Moreover, the gist can be used to guide people's choice of which specific objects to be recalled (Hollingworth & Henderson, 2000). For instance, when trying to retrieve the details of the scene, the gist can lead to recall of objects that are consistent with the scene, but were actually not present at all in the memory display (Lampinen, Copeland, & Neuschatz, 2001; Miller & Gazzaniga, 1998). Conversely, gist representations seem to facilitate the encoding of (semantic) outlier objects: items are more likely to be both fixated and encoded into memory when these are semantically inconsistent with the background scene (Hollingworth & Henderson, 2000, 2003). Arguably, these findings from studies with naturalistic scenes show that observers have a strong tendency to structure and organize a given sensory input into some higher-order regularity, that is, "compression" of the available information in order to spare the limited cognitive resources. These results seem to be strongly linked to studies that investigate the relational, or, hierarchical structure in objects (e.g., a global triangle composed of local squares, see Kimchi & Palmer, 1982). For instance, the identification of local-level elements in a hierarchical stimulus configuration (e.g., Navon letters) is influenced by representations at the global object level (Navon, 1977; Wagemans, Elder, et al., 2012). Moreover, global levels of a target object guide attention more efficiently during visual search than local object levels, and this global precedence in selecting a target on a given trial is transferred to subsequent trials, evidencing a persistent global bias (Conci, Müller, & Elliott, 2007a, 2007b; Conci, Töllner, Leszczynski, & Müller, 2011; Nie, Müller, Conci, 2016; Wagemans, Feldman, et al., 2012). Thus, an observer's representation of both real-world scenes and simpler displays with geometric objects consists not only of information about the individual objects but also includes structural information and a broad, gist-like representation of the overall information presented.

In fact, it has been shown that perceptual organization also plays a significant role in vWM – even for rather simple memory arrays. For instance, when separate objects are grouped together into perceptual units (e.g., by means of closure or repetition), this also results in better vWM performance, as each unit in the group can be encoded into a perceptual Gestalt, thus improving memory capacity (Woodman, Vecera, & Luck, 2003; Xu, 2006; Xu & Chun, 2007). Moreover, maximizing the symmetry of an object via completion improves vWM performance (Chen, Müller, & Conci, 2016). Together, these findings point to the use of organizational principles to optimize the storage of items, so as to relieve vWM capacity (see also Jiang, Olson, & Chun, 2000).

Relatedly, there is mounting psychophysical evidence that even in simple memory displays, items are not treated independently (see Brady, Konkle, & Alvarez, 2011, for a review). For instance, if a display is changed from mostly dark squares to mostly bright squares, then observers notice this change more efficiently than a matched change that does not alter the global statistics of the scene (Alvarez & Oliva, 2009; Victor & Conte, 2004). Moreover, when computing the average visual representation in simple arrays of items from a given display, observers discount outlier objects to only represent the majority of consistent items (Haberman & Whitney, 2010). Brady and Alvarez (2011) reported further evidence suggesting that the representation of "ensemble statistics" influences the representation of individual items: Observers are biased in reporting the size of an individual item by the mean size of all (or of potentially task-relevant) items in a particular display – which they interpreted as reflecting the integration of information about the ensemble size of items in the display with information about the size of a particular item. However, existing formal models of the architecture and capacity of vWM do not take into account the possibility of such hierarchically structured representations, but only consider how many individual items are remembered when treated independently (Luck & Vogel, 2013; Ma, Husain, & Bays, 2014).

In the present study, we developed a hierarchical variant of the change detection task to investigate how different object levels (i.e., global or local representations) are represented in vWM. Within each trial, multiple hierarchical shapes were presented in a memory array, followed by a test probe that appeared after a brief delay. Observers were required to memorize all objects and hierarchical levels, and to indicate whether a change occurred in the probe item, irrespective of the level (global or local) where the change had occurred. In addition, we manipulated between-object repetition at both hierarchical levels, systematically varying