



## Original Articles

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



Qi-Yang Nie<sup>a,\*</sup>, Hermann J. Müller<sup>a,b</sup>, Markus Conci<sup>a</sup>

<sup>a</sup> Department Psychologie, Ludwig-Maximilians-Universität München, Munich, Germany

<sup>b</sup> 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

## 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.

© 2016 Elsevier B.V. All rights reserved.

## 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](mailto:qiyang.nie@psy.lmu.de) (Q.-Y. Nie).

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

the repetition in displays with several hierarchical objects presenting similar objects at global or local levels. These variations of repetition permitted investigating how repetitions of object identities within a given display potentially affect the summary representation that is generated alongside with individual item memory. Both hierarchical structure and object repetition are known to provide a structural representation and corresponding statistical information about the objects that are to be remembered (Alvarez, 2011; Kimchi & Palmer, 1982).

To anticipate, our results demonstrate a form of hierarchical storage in vWM: The remembered object representation of individual items was biased toward the global object level, with better memory performance in detecting global as compared to local changes. Moreover, object repetitions also affected vWM capacity – with a reduction in performance that was evident in particular for globally repeated objects in a given display. This suggests that, contrary to existing models of vWM, items are not retrieved as integrated and independent units, but instead, an item's (reported) representation is constructed by combining global and local structure about that specific item with information about the set of items at a global level of the available ensemble statistics.

## 2. Experiments 1A and 1B

Experiments 1A and 1B were performed to investigate the representation of object structure in visual working memory (vWM) using a variant of the change detection task with hierarchical shapes. Observers were presented with a memory array that contained varying numbers of shapes with a global/local structure. Subsequent to a delay, a test probe was presented that could either undergo a change at a global level, at a local level, or remain the same (at both levels) compared to the target item at the same location in the preceding memory array (see Fig. 1 for examples; see

also Kimchi & Palmer, 1982). The key difference between the two experiments was the manipulation of set size, that is, the number of items that participants were required to remember. Set sizes were 2, 4, and 6 in Experiment 1A, and 1, 2, and 3 in Experiment 1B. On the basis of previous findings from visual search studies (e.g., Conci et al., 2007a, 2007b; Nie, Maurer, Müller, & Conci, 2016), we expected greater detection sensitivity and higher memory capacity for global as compared to local object representations.

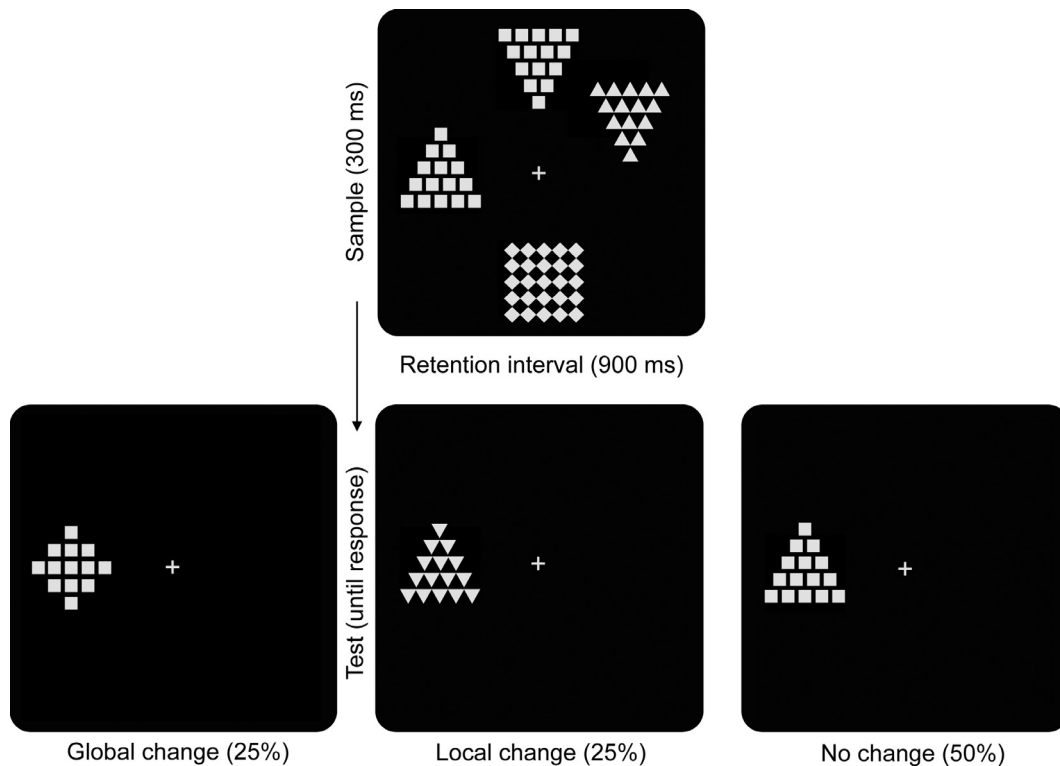
### 2.1. Methods

#### 2.1.1. Participants

Two different groups of observers participated in Experiments 1A (N = 12; 8 female; age range from 19 to 32 years; mean age = 20.5 years) and 1B (N = 10; 7 female; age range from 20 to 31 years; mean age = 22.8 years). All participants reported normal or corrected-to-normal visual acuity. Participants provided written consent to the procedure of the experiment, which was approved by the local ethics committee, in accordance with the Declaration of Helsinki. They received course credits or payment of 8 Euro per hour for their participation.

#### 2.1.2. Apparatus and stimuli

The experiments were conducted with an IBM-PC compatible computer using Matlab routines and Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were hierarchical shapes (as in Kimchi & Palmer, 1982) presented in gray (8.5 cd/m<sup>2</sup>) against a black (0.02 cd/m<sup>2</sup>) background on a 17-in. monitor screen (1024 × 768 pixels). Each stimulus consisted of a global shape that subtended 2.6° × 2.6° of visual angle. Global shapes were constructed from various (13–25) identical local shapes, arranged in an invisible 5 × 5 grid. Local shapes covered an area



**Fig. 1.** Example trial sequence in Experiments 1A and 1B. Observers viewed a sample display that consisted of a variable number of to-be-remembered hierarchical shapes arranged in a circle (top panel). After a brief (blank) delay, a test display was presented that either presented a probe item with a change at the global object-level (left panel), a change at the local object-level (middle panel), or an unchanged item (right panel).

of  $0.4^\circ \times 0.4^\circ$ . There was a  $0.15^\circ$  gap between each neighboring local shape.

Memory arrays consisted of 2, 4, and 6 hierarchical shapes in Experiment 1A, and of 1, 2, and 3 shapes in Experiment 1B. All shapes were presented on an imaginary circle of  $6^\circ$  radius, with positions randomly selected from eight equally spaced, fixed locations on the circle. Each hierarchical shape was constructed randomly from a predefined set of 4 distinctive local shapes (squares, diamonds, and up- or downward-pointing triangles), thus forming 12 different shape configurations, with the constraint that global and local shapes were always different from each other. Target (probe) locations were randomly selected on each trial. Fig. 1 shows an example display with set size 4, and three possible variants of test probes, which illustrate global and local changes, and the no-change condition (left, middle, and right panels, respectively).

### 2.1.3. Trial sequence

Each trial started with the presentation of a central fixation cross for 500 ms. The fixation cross was followed by the memory display presented for 300 ms; observers were instructed to memorize all presented hierarchical shapes in this (memory) display at both global and local levels. Subsequent to the memory display, there was a blank screen for 900 ms, followed by a test probe that was presented at one, randomly chosen location from the preceding memory array. The task was to decide whether the test probe was the same (at both global and local levels) or different (with a change at either the global or local level) relative to the item that had been previously presented at the same location in the memory array. The probe item remained on-screen until a response was recorded. Participants were instructed to respond as accurately as possible without emphasizing response speed. In case of an erroneous response, feedback was provided by an alerting red sign (“-”) presented for 1000 ms at the center of the screen. Each trial was separated from the next by a 500-ms interval.

### 2.1.4. Design and procedure

A three-factor within-subjects design was used for both experiments. The independent variables were change, level, and set size. The first variable, change indicated the memory-probe transition and could be either present (50% of the trials) or absent (50%). The second variable, level, refers to the hierarchical level at which a potential change occurred (global or local). For a global change, the test probe differed from the memorized target item (at the same location) only at the global level (Fig. 1, lower left), whereas for a local change, the test probe differed from the target item only at the local level (Fig. 1, lower middle). The third variable, set size, had three levels and determined the number of items presented in the memory array: 2, 4, and 6 hierarchical shapes in Experiment 1A and 1, 2, and 3 in Experiment 1B, respectively.

At the beginning of both experiments, participants completed 1 block of 48 practice trials generated randomly to familiarize them with the task. The subsequent, actual experiment then presented 576 trials, divided into 12 blocks of 48 trials each.

### 2.1.5. Data analysis

In the present experiments (as well as the subsequent ones), vWM performance was determined by a signal-detection-theoretic sensitivity measure: d-prime (see Macmillan & Creelman, 2004). For Experiments 1A and 1B, we also estimated the number of individual objects remembered (at global or local levels) using Cowan’s K (Cowan, 2001):  $K = (H - FA) \times N$ , where K is the number of items stored, H is the hit rate, FA is the false alarm rate, and N is the number of items presented.

## 2.2. Results

Fig. 2 shows mean d-primes and corresponding Cowan’s K estimates as a function of set size, separately for global and local level changes. Results are combined across Experiments 1A (gray symbols) and 1B (white symbols).

### 2.2.1. Experiment 1A

D-prime. Fig. 2A (gray symbols) presents the mean d-primes as a function of set size, separately for global and local level changes. Individual d-primes were subjected to a  $2 \times 3$  repeated-measures analysis of variance (ANOVA) with the factors level (global, local) and set size (2, 4, and 6). This analysis revealed both main effects to be significant: level,  $F(1,11) = 62.1$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.85$ , and set size,  $F(2,22) = 97.03$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.9$ . Global changes were detected more efficiently than local changes (mean precedence effect in d-prime: 0.89), and d-primes decreased (by 1.5,  $p < 0.001$ ) as set size increased (from 2 to 6). The two-way interaction was not significant,  $F(2,22) = 2.17$ ,  $p = 0.14$ ,  $\eta_p^2 = 0.17$ : global changes yielded a comparable decrease in d-primes across set sizes as local changes. To summarize, change detection sensitivity decreased with increasing set size, and sensitivity was much greater overall for global than for local changes. In addition, the decrease in d-prime across set size was comparable for both object levels, with the global precedence in change detection being rather stable across the number of to-be-remembered items.

Cowan’s K. A subsequent analysis examined working memory capacity for global and local changes, that is, whether memory capacity for the global object level would be larger (i.e., indicative of “precedence”) compared to that for the local level. To this end, a

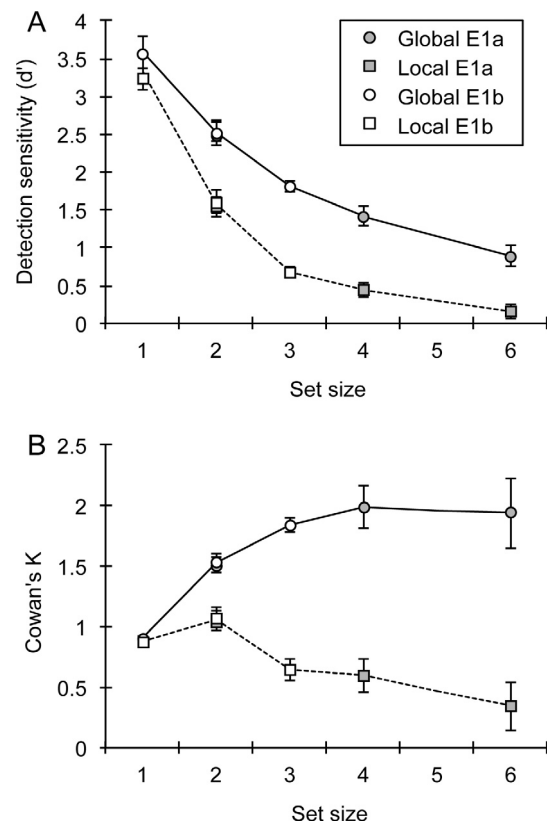


Fig. 2. Mean change detection sensitivity, d-prime (A) and corresponding capacity estimates, K (B) in Experiment 1, presented as a function of set size for global and local changes (solid and dashed lines, respectively). Error bars represent  $\pm 1$  SEM. Gray (set size: 2, 4, and 6) and white (set size: 1, 2, and 3) symbol colors represent values from Experiments 1A and 1B, respectively.



further  $2 \times 3$  repeated-measures ANOVA was performed on individual Cowan's  $K$  estimates with the factors level (global, local) and set size (2, 4, and 6). This analysis revealed a significant main effect of level,  $F(1,11) = 52.7$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.83$ , but no effect of set size,  $F(2,22) = 0.54$ ,  $p = 0.6$ ,  $\eta_p^2 = 0.05$ . The global object level was associated with a higher  $K$  estimate than the local level (1.97 vs. 0.47), with overall comparable  $K$  values across set size. The two-way interaction, was also significant,  $F(2,22) = 17.6$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.62$ . Post-hoc comparisons showed that memory precedence for the global object level increased with set sizes larger than 2 (mean precedence in  $K$ : 0.46, 1.4, and 1.59, respectively for set size 2, 4, and 6, all  $t(11) > 5$ ,  $ps < 0.001$ ). This result shows that, while the representation of the global object level has a greater memory capacity compared to the local level, the difference between levels becomes more pronounced with larger set sizes.

### 2.2.2. Experiment 1B

**D-prime.** In Experiment 1B, observers were presented with smaller set sizes (1, 2, and 3 items) of (global/local) hierarchical shapes. Despite this reduced set size, the results replicated those of Experiment 1A (see Fig. 2A). Fig. 2A (white symbols) presents the mean  $d$ -primes as a function of set size, separately for global and local changes. A  $2 \times 3$  repeated-measures ANOVA of the individual  $d$ -primes, with the factors level (global, local) and set size (1, 2, and 3), revealed both main effects to be significant: level,  $F(1,9) = 47.5$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.84$ , and set size,  $F(2,18) = 94.5$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.91$ . Detection sensitivity was again superior for global as compared to local changes (mean precedence effect in  $d$ -prime: 0.78), and sensitivity decreased (by 2.2) as set size increased from 1 to 3 (all  $ps < 0.001$ ). Moreover, the two-way interaction was significant,  $F(2,18) = 8.2$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.48$ : the benefit for global (relative to local) changes increased with larger memory arrays (mean precedence effects in  $d$ -prime were 0.27, 0.94, and 1.13 for set sizes 1, 2, and 3, respectively). To summarize, global change sensitivity was much bigger overall than local change sensitivity, and this difference increased with set size, again indicating that change detection operates more efficiently at the global object level.

**Cowan's  $K$ .** Examination of working memory capacity for global and local changes, by means of a level (global, local)  $\times$  set size (1, 2, 3) ANOVA on the individual Cowan's  $K$  estimates, revealed both main effects to be significant: level,  $F(1,9) = 65.04$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.88$ , and set size,  $F(2,18) = 43.6$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.83$ . Capacity was higher for global than for local object levels ( $K$ s of 1.84 and 0.64, respectively), and it increased (by 0.35) along larger set size. Moreover, the two-way interaction was significant,  $F(2,18) = 61.7$ ,  $p < 0.008$ ,  $\eta_p^2 = 0.87$ : the global precedence effect was reliable only for set sizes 2 and 3 (0.49 and 1.2, respectively;  $t(9)s > 4.6$ ,  $ps \leq 0.001$ ), but not for set size 1 (0.04,  $t(9) < 1$ ,  $p > 0.2$ ; see Fig. 2B). In other words, with single objects, observers were capable of remembering both levels to a comparable extent; but, as set size increased from 2 to 3, global object representations were more reliable than local representations (with the local level in fact exhibiting a reduction of memory capacity from 2 to 3 to-be-remembered objects).

### 2.3. Discussion

Experiment 1 examined how the hierarchical (global/local) representation of objects affects vWM performance. Across both Experiments 1A and 1B, change detection performance was found to be superior (in terms of  $d$ -prime and  $K$  scores) at the global object level compared to the local level. Also, despite a general decrease in performance with increasing memory load, differential performance between object levels only emerged with larger set

sizes: while change detection of a single object could be performed with comparable efficiency at both global and local levels, from set size 2 onwards, the global level showed a reliable benefit compared to the local level.

Analysis of the corresponding vWM capacity estimates revealed that observers could represent up to about two objects at the global level (see Fig. 2B, which shows an asymptote at  $K = 2$  for the global level). At the local level, by contrast, observers could only represent up to one object; and, in fact, with larger set sizes (i.e., 3 or more objects), the  $K$  estimate for the local level is numerically reduced (see Fig. 2B). Thus, when combining  $K$  values for the global and local levels, our results indicate that the memory capacity for hierarchical shapes taken together is at about 2.5 objects.

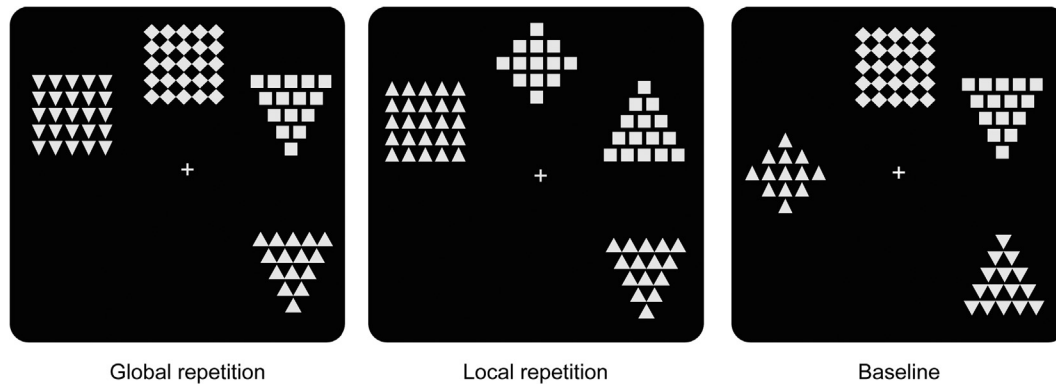
## 3. Experiment 2

Experiment 1 yielded a robust global precedence effect in vWM, which was evident both in measures of  $d$ -prime and in terms of memory capacity  $K$  – suggesting that the structure of to-be-remembered objects biases memory storage toward global object levels. Such a global bias might be related to the initial analysis of scenes in terms of their overall “gist” (Hollingworth & Henderson, 2003; Oliva, 2005; see also Conci & Müller, 2014). This representation of scene gist might in turn be related to summary statistics that represent sets of objects as a group or ensemble (e.g., of the average size of items; Alvarez, 2011; Brady & Alvarez, 2011). In other words, the global bias in vWM might, to some extent, be related to the analysis of global scene properties, which provides a summary representation of a given memory array.

To examine how such processes of scene analysis are related to global and local levels, in Experiment 2, we performed a change detection experiment with hierarchical shapes (essentially as in Experiment 1) but now varying the repetitions between to-be-remembered objects. Note that feature similarity (e.g., different degrees of redness) has been shown to affect object representations in vWM (e.g., Lin & Luck, 2008). Experiment 2 always presented a set size of four objects, and repetitions among items were manipulated systematically at the global and local levels: In the global repetition condition, always two (of the four) global-level objects were identical, and were presented alongside with four distinctive local shapes (Fig. 4, left panel). By contrast, in the local repetition condition, two local-level objects were always identical and were presented with four distinctive global shapes (Fig. 4, middle panel). Finally, a baseline condition presented four different hierarchical objects at both global and local levels (Fig. 4, right panel). Accordingly, the manipulation of global versus local repetition allowed us to examine the relative influence of processing and resolving global and local levels of memorized objects.

### 3.1. Methods

Experiment 2 was essentially identical to the previous experiments, except for a fixed set size of (always) 4 hierarchical shapes. Moreover, similarities between items were manipulated at three different levels: (i) global repetition, (ii) local repetition, and (iii) baseline. In the global and local repetition conditions, always two pairs of repeated global or local objects were presented together with four distinctive objects at the respectively other (local or global) level. In the baseline condition, there were four distinctive objects at both levels (see Fig. 3 for examples). All three types of display (global repetition, local repetition, and baseline) were presented with equal probability in both change and no-change conditions. Participants initially completed 1 block of 48 practice trials to become familiar with the task. The formal experiment was divided into 12 blocks of 48 trials each. Twelve new observers

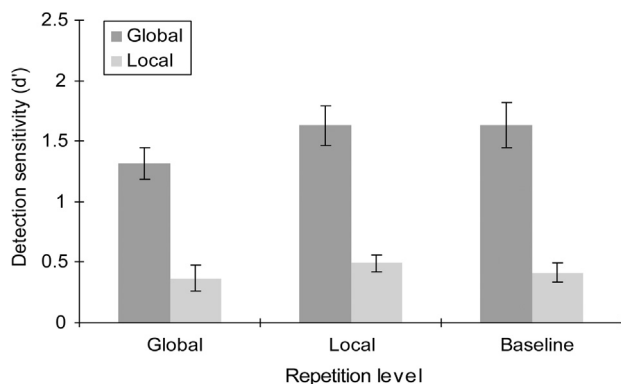


**Fig. 3.** Examples of memory arrays with four hierarchical objects as presented in Experiment 2. In the global repetition condition (left), two pairs of globally repeated objects were presented, while the local shapes were all distinctive. The local repetition condition presented two pairs of locally repeated objects, while in turn the global shapes were all different from each other (middle). In the baseline condition (right), all objects were different from each other at both the global and local levels.

(9 female; age from 21 to 31 years; mean age = 24.9 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour.

### 3.2. Results

Fig. 4 presents the mean  $d'$ -primes as a function of the repetition level, separately for global and local object changes. Individual  $d'$ -primes were subjected to a  $2 \times 3$  repeated-measures ANOVA with the factors level (global, local) and repetition (global, local, and baseline). This analysis revealed both main effects to be significant: level,  $F(1,11) = 48.2$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.81$ , and repetition,  $F(2,22) = 4.02$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.27$ . Global changes were detected more efficiently than local changes (mean precedence effect in  $d'$ -prime: 1.1), comparable to the findings in Experiment 1. Moreover, post hoc comparisons to decompose the main effect of repetition revealed that the mean  $d'$ -prime (averaged over global- and local-change conditions) was reduced when comparing object repetitions for global relative to baseline and global relative to local conditions (mean  $d'$ -prime differences were 0.18 and 0.22, respectively,  $ps < 0.04$ ). By contrast, there was no difference in memory performance when comparing local repetition with baseline ( $p = 0.99$ ). The two-way interaction was not significant,  $F(2,22) = 2.5$ ,  $p = 0.11$ ,  $\eta_p^2 = 0.19$ , indicating that global object repetition modulated detection of changes to a comparable extent at both global and local levels.



**Fig. 4.** Mean change detection sensitivity ( $d'$ -prime) in Experiment 2, presented as a function of the repetition level for both global and local changes. Error bars represent  $\pm 1$  SEM.

### 3.3. Discussion

Experiment 2 replicated Experiment 1 in showing an overall global bias in vWM. In addition, the results revealed an increase in object repetition at the global level to reduce mnemonic performance (affecting changes at both object levels in a similar manner). With multiple different to-be-memorized hierarchical shapes, observers' reports of a change of a given hierarchical shape displayed a cost deriving from the globally repeated objects in the memory array.

Our finding is compatible with the idea that vWM capacity reduces with competition between similar representations (Luck & Vogel, 2013). This notion would predict capacity (or, representational precision) to be lower when the to-be-remembered items are similar to each other. However, some of the available evidence appears to suggest that similarity-based perceptual grouping results in more accurate mnemonic representations, thus facilitating memory performance (Lin & Luck, 2008). It is possible that ensembles are represented in two different formats: either as ensemble averages, representing the average feature of to-be-remembered objects (as in simple feature estimation, see Bronfman, Brezis, Jacobson, & Usher, 2014); or, alternatively, as ensemble repetitions, representing the common features of presented objects, which is evident with high-order object representations in natural scenes. Accordingly, an ensemble may have distinct influences on the representation in vWM. For similar colored squares (Lin & Luck, 2008), ensemble averages of homogenous colors will lead to more precise mnemonic representation, whereas for globally repeated objects in the current experiment, ensemble repetitions likely provide an overall coarse representation of the entire display (Cohen, Dennett, & Kanwisher, 2016). Thus, our results suggest that ensemble repetitions of hierarchical memory representations interfere with mnemonic performance for individual objects. A potential alternative (not mutually exclusive) account for the reduction in performance with global object repetitions might be that observers were more likely to confuse, or "misbind" the memorized items when these were globally repeated (e.g., Bays, Wu, & Husain, 2011, for illusory bindings in vWM), illustrating once again the special role of the global scene layout in vWM, which would still be consistent with the account of global ensemble coding.

## 4. Experiment 3

Experiments 1 and 2 suggest that object structure influences vWM, such that more accurate information is retained from the global (relative to a local) object representation. One potential

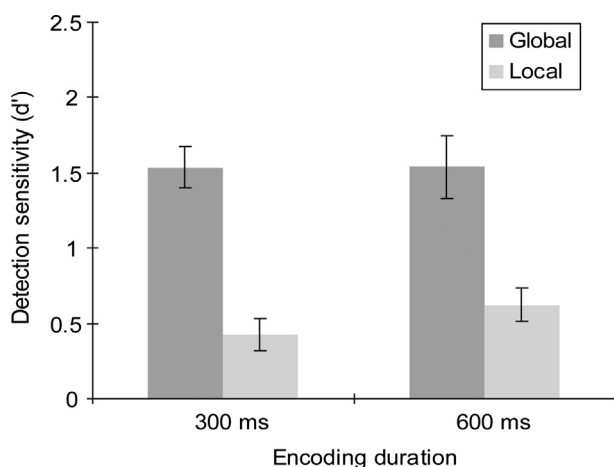
reason for this difference in vWM performance, which results in an advantage for global relative to local object levels, could simply be related to stimulus encoding, that is, processes reflecting basic perceptual processing. Therefore, to ensure that observers had sufficient time to encode the stimuli presented, in Experiment 3, we varied the duration of the sample array, comparing a 300-ms presentation time of the sample display (as used in Experiments 1 and 2) with a longer presentation duration of 600 ms. Note that both presentation durations are consistent with typical encoding durations in standard vWM experiments (see Luck & Vogel, 2013, for a review), whereas a much shorter or longer presentation duration might additionally involve the recruitment of iconic memory or internal rehearsal of the memorized content (Baddeley, 1986), respectively. A previous estimate based on performance in a change detection task suggested that the rate of encoding objects into memory occurs at approximately 50 ms per item (Vogel, Woodman, & Luck, 2006). Thus, the longer stimulus duration provided substantially more time to perceptually encode the stimulus configurations, potentially leading to improved local-level detection performance if time were indeed a limiting factor.

#### 4.1. Methods

Experiment 3 presented the baseline condition of Experiment 2 (with a fixed set size of four items, presenting distinctive shapes at both global and local levels), but with two different encoding durations of the memory array (presenting either a 300-ms or a 600-ms display duration, randomly intermixed within blocks). Participants initially completed 1 block of 48 practice trials, followed by 384 experimental trials. The experiment was divided into 8 blocks of 48 trials each. Eleven new observers (8 female; age from 21 to 31 years; mean age = 22.5 years) with normal or corrected-to-normal visual acuity participated in the experiment, receiving course credits or payment of 8 Euro per hour.

#### 4.2. Results

Fig. 5 presents the mean  $d'$ -primes as a function of the encoding duration, separately for global- and local-level changes. Individual  $d'$ -primes were subjected to a  $2 \times 2$  repeated-measures ANOVA with the factors level (global, local) and encoding duration (300, 600 ms). This analysis revealed only a significant main effect of level:  $F(1,10) = 37.5$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.79$ . Global changes were detected more efficiently than local changes (mean precedence



**Fig. 5.** Mean change detection sensitivity ( $d'$ -prime) in Experiment 3, presenting global and local changes for short and long encoding durations. Error bars represent  $\pm 1$  SEM.

effect in  $d'$ -prime: 1.01). Importantly, however, there were no effects involving the factor encoding duration ( $ps > 0.14$ ), suggesting that global precedence in working memory is not due to encoding limitations that might arise because of a too short duration of the presented memory array.

#### 4.3. Discussion

We again replicated a robust global precedence effect in vWM as already seen in the previous experiments. Our results also demonstrate that performance was not significantly influenced by variations of the encoding duration: Reliable global precedence effects in vWM were found with both shorter (300 ms) and longer (600 ms) durations of the memory display. Following prolonged exposure of a stimulus array, mnemonic representations should be constrained by limits of storage only (Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011). Thus, our findings likely indicate that the global/local hierarchy primarily reflects limitations in storage capacity, rather than limitations in perceiving, that is, encoding of the presented objects.

### 5. Experiment 4

The present results demonstrate that structural relations in objects are represented in vWM, revealing a global precedence effect that might originate from limited mnemonic resources. However, a potential alternative explanation to account for these results might be that global object representations are to some extent more salient than corresponding local object representations. For instance, a global change comprises a change to a single, large (global) configuration while all local elements remain constant. Conversely, in the case of a local change, many, small (local) shapes undergo a change (and the global configuration remains the same). Thus, differences in the number of the depicted changes, differences in relative size and/or the amount of crowding between change levels, may provide potential confounds that could alternatively (at least to some extent) account for our findings (see also Kimchi, 1992; Navon, 1981).

Experiment 4 was performed to test whether the detection of changes at global and local object levels differs when the changes (at varying levels) occur independently of each other. To this end, Experiment 4 essentially repeated the 300-ms presentation duration condition of Experiment 3, except that global- and local-change detections were now presented in separate sessions, such that participants only needed to memorize one task-relevant object level while ignoring the other level in the respective session. If the global bias is predominantly related to vWM maintenance of multiple hierarchical levels, we would predict that global memory precedence is substantially reduced when only one specific object level is relevant, while any residual differences might be taken to reflect additional influences that relate to stimulus saliency (e.g., relative size, or crowding).

#### 5.1. Methods

Experiments 4 presented the 300-ms display duration condition of Experiment 3, but with global and local changes presented in separate halves of the experiment (counterbalanced across participants). A new group of eleven observers (5 female; age from 18 to 31 years; mean age = 20.4 years) performed the global and local change-detection tasks in two separate, consecutive sessions. Each session started with a practice block of 24 trials, followed by 96 experimental trials that were divided into 4 blocks of 24 trials each. All participants had normal or corrected-to-normal visual acuity, and received course credits or payment of 8 Euro per hour.

## 5.2. Results

Fig. 6A presents the  $d'$ -prime values separately for global and local change detections. Paired  $t$  test revealed that  $d'$ -primes for detecting a global change were significantly higher than local change detections: 3.06 vs. 2.52,  $t(10) = 5.48$ ,  $p < 0.001$ , indicating that the global object level is still processed with priority even when only one hierarchical level is task-relevant during an entire half of the experiment.

In a next step, individual  $d'$ -primes were subjected to a  $2 \times 2$  mixed-design ANOVA with the within-subject factor level (global, local) and the between-subject factor block type (mixed [Experiment 3], separate [Experiment 4]). In the mixed block type, global and local change trials were presented in *mixed* order within a given block of trials (Experiment 3, 300-ms encoding duration), while for the *separate* block type global and local changes were presented in separate blocks (Experiment 4). The results from this analysis revealed significant main effects of level:  $F(1,20) = 73.1$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.78$  and of block type:  $F(1,20) = 57.1$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.74$ . The main effect of level simply depicts the above mentioned global precedence effect (i.e., the difference in  $d'$ -prime between global and local change detections), which was present with both mixed and separate presentations. Moreover, mixed, relative to separate, block types led to substantially reduced change

detection sensitivities ( $d'$ -primes: 0.98 vs. 2.8, see Figs. 5 and 6A, respectively). Importantly though, our analysis also revealed a significant level by block type interaction,  $F(1,20) = 8.6$ ,  $p < 0.009$ ,  $\eta_p^2 = 0.30$ , showing that the global precedence effect in the current experiment (*separate* block type) is significantly reduced as compared to Experiment 3 (*mixed* block type; global precedence in  $d'$ -primes: 0.54 vs. 1.11, respectively; see Fig. 6B). This finding suggests that global precedence is substantially reduced with the number of task-relevant object levels.

As described above, monitoring only one task-relevant object level in Experiment 4 (as opposed to both object levels in previous experiments) led to a substantial increase in the overall detection sensitivity. However, this overall difference in performance might additionally have influenced the size of the global precedence effect. We therefore calculated a *relative* global precedence score that normalizes the difference between global and local change detections relative to the default, global level of performance (i.e.,  $[\text{global} - \text{local}]/\text{global } d'$ -primes). This relative difference score revealed that global precedence in Experiment 4 modulated detection performance by 17.6%, as compared to a much larger difference of 72.1% in Experiment 3.

## 5.3. Discussion

Without having to remember both global and local object levels simultaneously, the performance for both global- and local-change detections was found to be significantly enhanced. Nevertheless, we still obtained a reliable (but reduced) global precedence effect that showed a bias in processing global-level object information. This suggests that a small, but yet reliable part of the global precedence effect might be attributed to differences in saliency between global and local levels of representation (relative global precedence effect of 17.6%). Importantly, however, the larger part of the global precedence effect (i.e., 72.1%) appears to be related to hierarchical-level differences as they are maintained in vWM.

## 6. General discussion

Studies that investigate vWM typically present relatively simple features or objects that are assumed to be represented independently of each other. However, recent evidence suggests that, rather than representing only individual items, vWM also provides a structural representation, that is: ensemble statistics relating to aspects of the “gist” of the presented scene (Alvarez, 2011; Brady & Alvarez, 2011; Brady et al., 2011). In an attempt to elaborate this notion, the aim of the present study was to investigate how hierarchical relations within and across objects are represented in vWM. Our experiments yielded four main results, namely: (i) vWM representations are organized in a hierarchically structured fashion reflecting the global/local structure of the perceptual input; (ii) repetition between items particularly at a global object level gives rise to vWM capacity detriments; (iii) global precedence effects in the current change detection paradigm primarily reflect the way items are stored during the retention phase; and (iv) this global benefit mainly reflects the globality of memory itself, and can only be partially attributed to differences in saliency across object levels.

### 6.1. Beyond objects and features: hierarchical representations in vWM

To explore the hierarchical structure within given object representations in vWM, we introduced a change detection task with hierarchical, global/local shapes, in which a change occurred either at the global or the local level of a given target object. We found a robust pattern of global precedence: Measures of memory

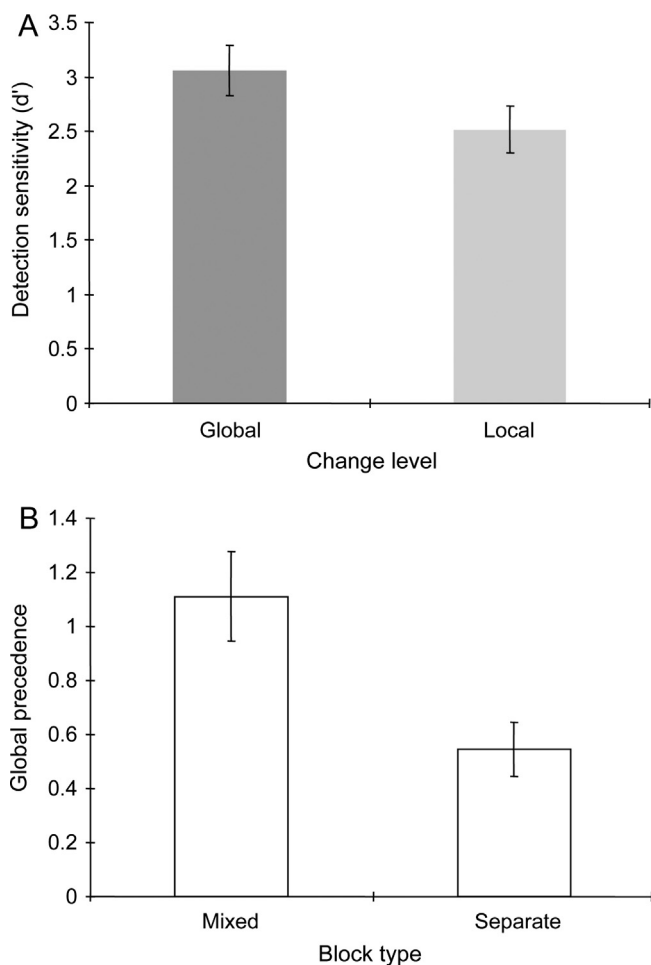


Fig. 6. Mean change detection sensitivity,  $d'$ -prime (A) in Experiment 4, with global and local changes presented in separate halves of the experiment. (B) Global precedence effect (global minus local  $d'$ -primes) in Experiment 3 (300-ms encoding duration, in which global and local change trial were *mixed* within blocks) and in Experiment 4 (with global and local change trials presented in *separate* blocks). Error bars represent  $\pm 1$  SEM.



performance (d-prime, Cowan's K) for the global object level revealed higher sensitivity and larger capacity compared to measures for the local level. Importantly, this global memory bias increased significantly with larger set sizes, reaching an asymptote at a K value of approximately 2; the corresponding local memory capacity was overall smaller, with a K value of about 1 that decreased to 0.5 with larger memory arrays. Thus, limited memory resources are distributed asymmetrically across object levels: with an increased memory load, the global benefit becomes larger. Moreover, the total amount of visual information maintained in vWM turned out to be rather stable, with a maximum capacity of around  $K=2.5$  items when global and local levels are pooled together. This is consistent with the view that, overall, there is a limited amount of visual information that can be held in vWM (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Ma et al., 2014). Our study extends this idea by showing that the distribution of limited memory resources is hierarchically organized within a given object, reflecting the inherent, to-be-remembered visual structure.

Any measure of memory capacity is meant to estimate its underlying "units", where what exactly counts as the proper unit depends on the structure of the represented information. For instance, it has been suggested that objects are represented as separate visual features that are stored in independent "channels", each with their own capacity limitation (e.g., color or orientation; Magnussen, Greenlee, & Thomas, 1996), or in terms of integrated object representations (Luck & Vogel, 1997; Vogel et al., 2001). Recent evidence suggests that there are significant benefits to remembering multiple features of a single object compared to the same set of features distributed across multiple objects (Fougnie, Asplund, & Marois, 2010; Olson & Jiang, 2002). For instance, it is easier to store five different colors and five orientations that define the same five objects than to store the same 10 features for separate objects (Fougnie, Cormiea, & Alvarez, 2013). The finding that vWM improves with fewer discrete objects (Olson & Jiang, 2002) has been taken to suggest that the representations that underlie vWM are object-based (Luck & Vogel, 1997; Vogel et al., 2001). According to this notion, vWM can store a small, fixed number of objects and integrate multiple features into a single object representation. However, there is also evidence showing that vWM representations are, in fact, not purely object-based. For instance, having to remember more features engenders significant costs in terms of the fidelity of each feature representation (Fougnie et al., 2010). Moreover, attending to one relevant feature reduces the mnemonic precision of a second task-irrelevant feature of the same object, such that memory precision for multiple features of an object may vary as a function of the attentional engagement devoted to each feature (Shin & Ma, 2016; Swan, Collins, & Wyble, 2016). These results indicate that what counts as the right "unit" in vWM is neither a fully integrated object representation nor a collection of independent features. Instead, vWM units appear to encompass rather flexible organizational principles. In light of the current results, the actual unit in vWM would appear to reflect a hierarchical structure, with the global-level representations of this "unit" being prioritized for vWM storage, while less resources are assigned to local-level object representations.

### 6.2. Beyond slots vs. resources: hierarchical ensembles in vWM

To further investigate the relational structure between object representations, we tested the role of inter-item repetition in Experiment 2, in which pairs of hierarchical objects were similar either at the global or at a local level, compared to a baseline condition in which none of the objects were similar to other items at the global/local object levels. While replicating the basic pattern established in Experiment 1, we further observed that repetition

at the global level interfered with the detection of both global and local changes. Local repetition, by contrast, appeared to have no influence (with performance comparable to the baseline level).

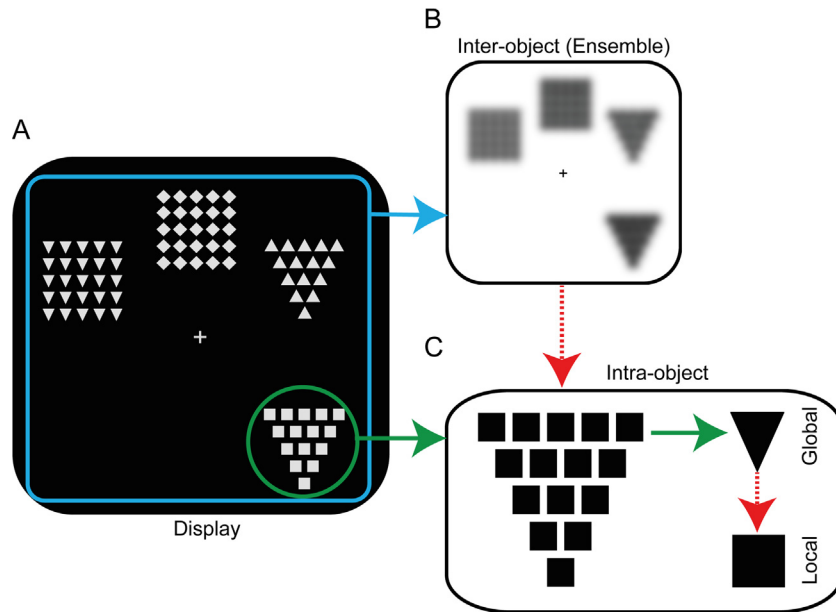
Prominent theories of vWM have proposed that memory limitations arise entirely from the availability of some limited resource that is either quantized into slots (Zhang & Luck, 2008) or continuously divisible (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Wilken & Ma, 2004). These notions, and the supporting studies, leave open the question of how contextual or ensemble representations interact with representations of individual items in vWM. A potential account in this regard assumes that the representation of ensemble statistics could take up space in memory that could otherwise be used to represent information about individual items (Cohen et al., 2016). In agreement with this view, our findings show that vWM representations are biased toward the global level and interference arises in particular among similar global-object representations. Restated, a global ensemble representation of the entire display impairs (via repetition) both global and local memory representations of individual items, illustrating that the different hierarchical object levels of vWM representations are linked and dependent on each other, rather than being maintained independently (Fougnie et al., 2013).

### 6.3. A theoretical model of hierarchical working memory

Recently, a hierarchical feature-bundle model has been proposed with the aim to integrate both object- and feature-based effects in vWM (Brady et al., 2011). According to this model, each unit of vWM is a hierarchically structured feature bundle, consisting of an integrated object representation at the top level and individual features represented at a lower level. The main idea of the model is that a unit in vWM is determined by the top level representation of an integrated object, while, at the same time, the lower level elemental feature of an object can be accessed by means of top-to-bottom decomposition within a given bundle.

In light of the current findings and in general agreement with the feature-bundle model, we propose that various levels of representation of a set of items in vWM are likewise stored in a hierarchically organized fashion. In a process paralleling how people memorize real scenes (Oliva, 2005), observers might represent summary statistics (i.e., the "gist") of the entire memory display in addition to information about each specific item. Each item in turn has its own global and local representation, which are also maintained in terms of a hierarchically structured representation with global and local object levels. This hierarchical storage format within and across individual items would permit observers to represent not only the individual identity of the to-be-remembered items, but also the structural relations (global/local) across the display layout. However, storing the various object levels plus the global ensemble might come at an overall cost, which is reflected in the overall low vWM capacity of about 2.5 items (as compared to capacity estimates of about 4 with displays that present relatively simple colored squares; e.g., Luck & Vogel, 1997).

A schematic model to accommodate the various levels of representation is depicted in Fig. 7 (on the basis of Brady et al., 2011). At an intra-object level, the vWM representation of a given individual item consists of two hierarchically structured layers, with a global object representation being stored at the top level and the corresponding local representation at the bottom level (Fig. 7C). In this view, representational units in vWM are conceived as being hierarchically structured across global/local levels, thus reflecting the global and local object properties of the stimulus input. The two layers of the representation are not equal; rather, the global level receives a representational bias. Consequently, the global object representation would be available at the top-level unit, whereas information concerning the local object would be stored at



**Fig. 7.** Example memory array (A) and a schematic model of hierarchically structured representations in vWM, designed to illustrate the interaction between inter-object ensemble representations (B) and storage of individual items and their intra-object (global/local) relations (C). At an inter-object level (blue arrows), the global display characteristics are encoded into an ensemble representation (B). In addition, individual items are stored at an intra-object level (C), with separate representations of the global and local levels of a given to-be-remembered object (green arrows). The model also incorporates inhibitory links between different levels (red, dashed arrows), reflecting the reduction of mnemonic precision from the inter- to the intra-object level (e.g., for multiple, repeated items), and from the global to local mnemonic representations, to account for our finding of a robust global precedence in vWM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the lower hierarchical level. Critically, the top-level unit receives priority such that the majority of the available mnemonic resources are used to maintain the global-level representation. By contrast, the subordinate, lower-level representation of local object information would receive only a smaller amount, that is, the remaining resources (as indicated by the dashed arrow) – in line with the current observation of global precedence in change detection. Of note, in the current experiments, an asymmetry in performance primarily results from global and local levels reflecting the inherent hierarchical structure of a given object. However, comparable differences in processing can also be observed in non-hierarchical objects with multiple features where an asymmetry in mnemonic performance results from varying levels of attentional engagement (Shin & Ma, 2016; Swan et al., 2016).

Moving beyond individual vWM representations, at an inter-object level, the entire display layout is retained in particular with reference to the global object characteristics; as a result, globally similar items are merged into a single ensemble representation (Fig. 7B). This ensemble representation essentially reflects the observed global bias overall, that is, global precedence and repetition effects might both arise from the global-level representation – enhancing the global objects but also causing interference as revealed by the impaired mnemonic representation of the entire individual object (dashed arrow).

One counterintuitive prediction of the model is that if the objects all had the exact same shape at the global level, performance should actually be worse than if there were several distinct global shapes presented. This is obviously not a likely experimental outcome as the task should be much easier when shapes at a given level are the same. A potential explanation might be that for homogeneous global display representations, redundant information presented at all item locations obviates the need to encode separate units of information from each location, but simply requires memorizing the repeated structure overall, thus reducing vWM

load. A potential constraint of our schematic model therefore is that it can only account for competition between several repeated global object representations in heterogeneous displays (as in the example of Fig. 7B), which leads to an overall impairment of individual (item) memory representation by means of ensemble coding.

In brief, this schematic model extends previous vWM models by taking into account the hierarchical relations both within and across objects, thus (to some extent) reflecting the typical structures in our natural environment with both representations of the overall scene layout and the more detailed object information. In this view, items are stored across three layers of representation, from the overall scene layout to the fine, detailed object information at the local level. Importantly, the model encompasses interference from top to bottom layers to illustrate the hierarchical organization of visual information, which in general assigns priority to the global level.

## 7. Conclusion

The present study reveals a functional connection between the representation of objects at varying hierarchical levels and the organization of vWM. Object representations in vWM are, by default, biased toward the global level, with the global bias existing across varying encoding durations and mainly reflecting the globality of memory itself. This suggests that global precedence in change detection primarily originates from hierarchically structured representations that are held in vWM. Memory performance is also influenced by the ensemble structure of the displays, that is: the interference of repetition among objects at the global level manifests in terms of impaired mnemonic representations for both global and local object levels. Together, our findings challenge models that propose that a fixed number of independent objects

can be remembered regardless of the presented object structure. Instead, our results support a more flexible account that emphasizes the role for hierarchically structured representations in vWM.

## Acknowledgments

This work was supported by project grants from the German Research Foundation (DFG; CO 1002/1-1), and from the “LMUexcellent” Junior Researcher Fund. We thank Xiaowei Ding for help with data collection in Experiment 4.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2016.11.009>.

## References

- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122–131. <http://dx.doi.org/10.1016/j.tics.2011.01.003>.
- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2), 106–111. <http://dx.doi.org/10.1111/j.0963-7214.2004.01502006.x>.
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, 106(18), 7345–7350. <http://dx.doi.org/10.1073/pnas.0808981106>.
- Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.
- Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, 11(10), 6. <http://dx.doi.org/10.1167/11.10.6>.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321, 851–854. <http://dx.doi.org/10.1126/science.1158023>.
- Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, 49(6), 1622–1631. <http://dx.doi.org/10.1016/j.neuropsychologia.2010.12.023>.
- Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, 22(3), 384–392. <http://dx.doi.org/10.1177/0956797610397956>.
- 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(5), 4. <http://dx.doi.org/10.1167/11.5.4>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. <http://dx.doi.org/10.1163/156856897X00357>.
- Bronfman, Z. Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: “Cost free” color phenomenality outside focal attention. *Psychological Science*, 25(7), 1394–1403. <http://dx.doi.org/10.1177/0956797614532656>.
- Chen, S., Müller, H. J., & Conci, M. (2016). Amodal completion in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1344–1353. <http://dx.doi.org/10.1037/xhp0000231>.
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, 20(5), 324–335. <http://dx.doi.org/10.1016/j.tics.2016.03.006>.
- Conci, M., & Müller, H. J. (2014). Global scene layout modulates contextual learning in change detection. *Frontiers in Psychology*, 5, 89. <http://dx.doi.org/10.3389/fpsyg.2014.00089>.
- Conci, M., Müller, H. J., & Elliott, M. A. (2007a). The contrasting impact of global and local object attributes on Kanizsa figure detection. *Perception & Psychophysics*, 69(8), 1278–1294. <http://dx.doi.org/10.3758/BF03192945>.
- Conci, M., Müller, H. J., & Elliott, M. A. (2007b). Closure of salient regions determines search for a collinear target configuration. *Perception & Psychophysics*, 69(1), 32–47. <http://dx.doi.org/10.3758/BF03194451>.
- Conci, M., Töllner, T., Leszczynski, M., & Müller, H. J. (2011). The time-course of global and local attentional guidance in Kanizsa-figure detection. *Neuropsychologia*, 49(9), 2456–2464. <http://dx.doi.org/10.1016/j.neuropsychologia.2011.04.023>.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114. <http://dx.doi.org/10.1017/S0140525X01003922>.
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, 10(12), 27. <http://dx.doi.org/10.1167/10.12.27>.
- Fougnie, D., Cormiea, S., & Alvarez, G. A. (2013). Object-based benefits without object-based representations. *Journal of Experimental Psychology: General*, 142(3), 621–626. <http://dx.doi.org/10.1037/a0030300>.
- Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, 72(7), 1825–1838. <http://dx.doi.org/10.3758/APP.72.7.1825>.
- Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition*, 7(1–3), 213–235. <http://dx.doi.org/10.1080/135062800394775>.
- Hollingworth, A., & Henderson, J. M. (2003). Testing a conceptual locus for the inconsistent object change detection advantage in real-world scenes. *Memory & Cognition*, 31(6), 930–940. <http://dx.doi.org/10.3758/BF03196446>.
- Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 683–702. <http://dx.doi.org/10.1037/0278-7393.26.3.683>.
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, 112(1), 24–38. <http://dx.doi.org/10.1037/0033-2909.112.1.24>.
- Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 8(4), 521–535. <http://dx.doi.org/10.1037/0096-1523.8.4.521>.
- Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5), 1211–1222. <http://dx.doi.org/10.1037/0278-7393.27.5.1211>.
- Lin, P.-H., & Luck, S. J. (2008). The influence of repetition on visual working memory representations. *Visual Cognition*, 17, 356–372. <http://dx.doi.org/10.1080/13506280701766313>.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281. <http://dx.doi.org/10.1038/36846>.
- Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, 17(8), 391–400. <http://dx.doi.org/10.1016/j.tics.2013.06.006>.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356. <http://dx.doi.org/10.1038/nn.3655>.
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Magnussen, S., Greenlee, M. W., & Thomas, J. P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, 22(1), 202–212. <http://dx.doi.org/10.1037/0096-1523.22.1.202>.
- Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, 36(6), 513–520. [http://dx.doi.org/10.1016/S0028-3932\(97\)00148-6](http://dx.doi.org/10.1016/S0028-3932(97)00148-6).
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3), 353–383. [http://dx.doi.org/10.1016/0010-0285\(77\)90012-3](http://dx.doi.org/10.1016/0010-0285(77)90012-3).
- Navon, D. (1981). The forest revisited: More on global precedence. *Psychological Research*, 43(1), 1–32. <http://dx.doi.org/10.1007/BF00309635>.
- Nie, Q.-Y., Müller, H. J., & Conci, M. (2016). Searching for forest or trees: Attentional zooming and level-specific memory in hierarchical objects. (submitted for publication).
- Nie, Q.-Y., Maurer, M., Müller, H. J., & Conci, M. (2016). Inhibition drives configural superiority of illusory Gestalt: Combined behavioral and drift-diffusion model evidence. *Cognition*, 150, 150–162. <http://dx.doi.org/10.1016/j.cognition.2016.02.007>.
- Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, 64(7), 1055–1067. <http://dx.doi.org/10.3758/BF03194756>.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <http://dx.doi.org/10.1163/156856897X00366>.
- Shin, H., & Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, 16(5), 1–8. <http://dx.doi.org/10.1167/16.5.10>.
- Swan, G., Collins, J., & Wyble, B. (2016). Memory for a single object has differently variable precisions for relevant and irrelevant features. *Journal of Vision*, 16(3), 1–12. <http://dx.doi.org/10.1167/16.3.32>.
- Victor, J. D., & Conte, M. M. (2004). Visual working memory for image statistics. *Vision Research*, 44(6), 541–556. <http://dx.doi.org/10.1016/j.visres.2003.11.001>.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92–114. <http://dx.doi.org/10.1037/0096-1523.27.1.92>.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1436–1451. <http://dx.doi.org/10.1037/0096-1523.32.6.1436>.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217. <http://dx.doi.org/10.1037/a0029333>.
- Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin*, 138(6), 1218–1252. <http://dx.doi.org/10.1037/a0029334>.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 11. <http://dx.doi.org/10.1167/4.12.11>.

- Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10(1), 80–87. <http://dx.doi.org/10.3758/BF03196470>.
- Xu, Y. (2006). Encoding objects in visual short-term memory: The roles of feature proximity and connectedness. *Perception & Psychophysics*, 68, 815–828. <http://dx.doi.org/10.3758/BF03193704>.
- Xu, Y., & Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proceedings of the National Academy of Sciences*, 104(47), 18766–18771. <http://dx.doi.org/10.1073/pnas.0705618104>.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235. <http://dx.doi.org/10.1038/nature06860>.