

# UNDERSTANDING THE OPERATION OF VISUAL WORKING MEMORY IN RICH COMPLEX VISUAL CONTEXT

EDITED BY: Hagit Magen, Marius Peelen, Tatiana Aloï Emmanouil and  
Zaifeng Gao  
PUBLISHED IN: Frontiers in Psychology





# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88966-104-6

DOI 10.3389/978-2-88966-104-6

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# UNDERSTANDING THE OPERATION OF VISUAL WORKING MEMORY IN RICH COMPLEX VISUAL CONTEXT

Topic Editors:

**Hagit Magen**, Hebrew University of Jerusalem, Israel

**Marius Peelen**, Radboud University Nijmegen, Netherlands

**Tatiana Aloï Emmanouil**, Baruch College (CUNY), United States

**Zaifeng Gao**, Zhejiang University, China

**Citation:** Magen, H., Peelen, M., Emmanouil, T. A., Gao, Z., eds. (2020).

Understanding the Operation of Visual Working Memory in Rich Complex Visual Context. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-104-6

# Table of Contents

- 04 Editorial: Understanding the Operation of Visual Working Memory in Rich Complex Visual Context**  
Hagit Magen, Marius V. Peelen, Tatiana Aloï Emmanouil and Zaifeng Gao
- 07 Relation Between Working Memory Capacity of Biological Movements and Fluid Intelligence**  
Tian Ye, Peng Li, Qiong Zhang, Quan Gu, Xiqian Lu, Zaifeng Gao and Mowei Shen
- 16 Visual Working Memory for Faces and Facial Expressions as a Useful “Tool” for Understanding Social and Affective Cognition**  
Filippo Gambarota and Paola Sessa
- 23 Persistence and Accumulation of Visual Memories for Objects in Scenes in 12-Month-Old Infants**  
Sylvia B. Guillory and Zsuzsa Kaldy
- 34 The Effect of Emotional Valence and Arousal on Visuo-Spatial Working Memory: Incidental Emotional Learning and Memory for Object-Location**  
Marco Costanzi, Beatrice Cianfanelli, Daniele Saraulli, Stefano Lasaponara, Fabrizio Doricchi, Vincenzo Cestari and Clelia Rossi-Arnaud
- 47 Composite Face Effect Predicts Configural Encoding in Visual Short-Term Memory**  
Lilian Azer and Weiwei Zhang
- 55 Spatial Organization in Self-Initiated Visual Working Memory**  
Hagit Magen and Tatiana Aloï Emmanouil
- 72 Feature Binding of Sequentially Presented Stimuli in Visual Working Memory**  
Anuj Kumar Bharti, Sandeep Kumar Yadav and Snehlata Jaswal
- 85 A Metacognitive Perspective of Visual Working Memory With Rich Complex Objects**  
Tomer Sahar, Yael Sidi and Tal Makovski
- 99 Visual Working Memory of Chinese Characters and Expertise: The Expert’s Memory Advantage is Based on Long-Term Knowledge of Visual Word Forms**  
Hubert D. Zimmer and Benjamin Fischer





# Editorial: Understanding the Operation of Visual Working Memory in Rich Complex Visual Context

Hagit Magen<sup>1\*</sup>, Marius V. Peelen<sup>2</sup>, Tatiana Aloï Emmanouil<sup>3</sup> and Zaifeng Gao<sup>4</sup>

<sup>1</sup> School of Occupational Therapy, The Hebrew University, Jerusalem, Israel, <sup>2</sup> Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, <sup>3</sup> Baruch College and the Graduate Center of the City University of New York, New York, NY, United States, <sup>4</sup> Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China

**Keywords:** visual working memory, gestalt, configuration, regularities, emotion, development, biological motion

## Editorial on the Research Topic

### Understanding the Operation of Visual Working Memory in Rich Complex Visual Context

As we move around in the world, we process a vast amount of visual information while we interact with objects or events, navigate our way in the environment, and process emotional and social information. Due to changes and interruptions in the visual input, for example during eye movements, task performance often relies on short-lived internal visual representations of the external world. Visual working memory (VWM) is the mechanism in charge of the formation and temporary maintenance of such representations (see Luck and Vogel, 2013; Ma et al., 2014, for reviews). Much work in the field of VWM has focused on identifying the basic units of VWM and the number of units that can be stored simultaneously, given its limited capacity (Cowan, 2001). Therefore, the majority of studies and theories in the field of VWM focus on the maintenance of basic, isolated, and often arbitrary visual stimuli, overlooking the involvement of VWM in the processing of highly rich and complex information, such as those frequently encountered in real life. The goal of the Research Topic was to broaden our knowledge regarding the maintenance of complex information in VWM. In the following sections, we provide a brief overview of the articles that appear in the collection. Among the issues addressed are: configural organization of VWM representations; metacognition for VWM performance; the role of emotions and of long-term memory (LTM) in VWM; and VWM for biological movements and real world objects.

Consistent with the structure present in natural environments, the basic representations of VWM are also structured, in a way that maintains not only individual item information but also configural properties of the entire display (Jiang et al., 2000; Brady et al., 2011). In this collection, Azer and Zhang adopted an individual differences approach to explore whether there is an association between configural processing in VWM and in perception. For each participant, the authors measured individual item and configural encoding in a VWM orientation task (Xie and Zhang, 2017) as well as holistic face processing using the composite face effect of the Le Grand face task (Le Grand et al., 2001). Configural encoding, but not individual item encoding, in VWM was correlated with holistic face processing. These findings support the hypothesis that perceiving configural properties in the environment leads to the encoding of these properties in VWM.

Most of the visual information we process is available simultaneously, affording the encoding of spatial configurations. However, information may also be accumulated over time, limiting the availability of spatial configural cues. Bharti et al. investigated the boundaries of configural processing in VWM by comparing the consequence of changing the locations of items between encoding and retrieval, when a set of binding stimuli was presented simultaneously or sequentially. In four experiments they consistently found that when a stimulus vanished as the next was

## OPEN ACCESS

### Edited and reviewed by:

Bernhard Hommel,  
Leiden University, Netherlands

### \*Correspondence:

Hagit Magen  
msmagen@mail.huji.ac.il

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 14 July 2020

**Accepted:** 20 July 2020

**Published:** 28 August 2020

### Citation:

Magen H, Peelen MV, Emmanouil TA  
and Gao Z (2020) Editorial:  
Understanding the Operation of Visual  
Working Memory in Rich Complex  
Visual Context.  
Front. Psychol. 11:1996.  
doi: 10.3389/fpsyg.2020.01996

presented, the WM advantage in the same-location condition over the randomized-location condition was dramatically reduced compared to the simultaneous presentation mode. These results suggest that WM encoding under the simultaneous presentation mode, but not under the sequential mode, relies on spatial configuration.

Magen and Emmanouil explored configural processing in self-initiated VWM, by examining the spatial structure of memory representations participants created for themselves, when space was task irrelevant. The results showed that participants constructed spatially structured representations for themselves and spatially unstructured representations for a hypothetical competitor in a memory contest. Nevertheless, in their explicit descriptions of the strategies they used, participants mentioned only non-spatial strategies. Thus, participants were guided by metacognitive knowledge on configural processing in VWM, knowledge that was implicit to some degree. The study shows that, mirroring the world, spatial structure is inherent to self-initiated VWM representations.

Metacognitive processing has gained little attention in VWM research. Sahar et al. explored how well individuals monitored their performance when maintaining real-world objects in VWM. Participants were tested independently on three dimensions: the objects' identity, location and temporal order. Monitoring was evaluated by comparing subjective confidence with objective performance. Similar biases in the subjective judgments of confidence were observed in all 3 dimensions, as reflected in overestimation of accuracy and underestimation of errors. Memory for real-world objects, which are represented in long-term memory (LTM), was enhanced relative to memory for meaningless images. Interestingly, monitoring of real-world objects was less biased.

Similarly investigating the influence of LTM on VWM, Zimmer and Fischer compared VWM for Chinese characters in Chinese and German participants. Across multiple experiments, they found that Chinese participants were better in detecting changes in the characters' shape but not in other aspects of these characters, such as their color or font type. These results provide novel evidence for an influence of LTM on VWM performance within the domain of word forms, contributing to a growing literature investigating the effects of familiarity and expertise on VWM performance (Jackson and Raymond, 2008; Curby et al., 2009; Kaiser et al., 2015; Xie and Zhang, 2017; Asp et al., 2019).

Only few studies investigated the developmental trajectory of VWM for complex materials. Guillory and Kaldy explored 12-month old infants' memory for real-world objects which were embedded in natural scenes. The authors examined whether, similar to adults, infants accumulate visual information in memory over time, and whether the formed memory representations are immune to interruptions. Infants' memory improved with longer exposures, showing accumulation of information over time. The formed representations were fragile as interference impaired memory performance. However, eye-tracking data demonstrated that some aspects of the scene

survived the interference. The study adds to our understanding of the development of VWM in ecological contexts.

Three articles in the collection explored the interaction and role of VWM in emotion and social contexts. Costanzi et al. tested whether memory for spatial locations that were associated with emotional information was enhanced relative to locations that were associated with neutral stimuli. Moreover, they systematically manipulated the level of valence and arousal of the emotional stimuli in four experiments. The results demonstrated that emotions enhance spatial WM performance when neutral and emotional stimuli compete with one another for access into VWM, as well as shedded light on the interplay between arousal and valence in driving information processing into VWM.

Gambarota and Sessa provided a comprehensive review on the representation of faces in VWM, through its interaction with LTM and emotional and social cognitive mechanisms. They reviewed studies comparing VWM representations of faces and of other classes of stimuli, summarized the findings on representing static and changeable facial features in VWM, and finally examined research showing qualitative differences in VWM for face representations as a function of psychopathology and personality traits. This review enabled us to have a panorama about the representation of faces in VWM and its potential role in supporting socio-affective cognition.

Ye et al. further explored the function of retaining biological movements in VWM. Biological movements are one of the most complex stimuli in our daily life, and contain rich social information. The study revealed that VWM capacity for biological movements not only predicts core social ability (Gao et al., 2016; He et al., 2019), but also predicts canonical cognitive ability (e.g., fluid intelligence).

Taken together, the studies in this topic provide a more ecological view of VWM, seen as an adaptive system that evolved to reflect the structure of natural regularities, prioritize social, and emotional information which is necessary for survival, and integrates long-term knowledge of the environment. The articles highlight some of the important topics that need to be studied further and be incorporated into a comprehensive, more ecological model of VWM.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

## FUNDING

MP received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 725970). ZG received funding from the National Natural Science Foundation of China (31771202).

## REFERENCES

- Asp, I., Störmer, V. S., and Brady, T. F. (2019). Greater visual working memory capacity for visually-matched stimuli when they are recognized as meaningful. *PsyArXiv [preprint]*. doi: 10.31234/osf.io/r6njf
- Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vision* 11, 1–34. doi: 10.1167/11.5.4
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. doi: 10.1017/S0140525X01003922
- Curby, K. M., Glazek, K., and Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 94–107. doi: 10.1037/0096-1523.35.1.94
- Gao, Z., Ye, T., Shen, M., and Perry, A. (2016). Working memory capacity of biological movements predicts empathy traits. *Psychon. Bull. Rev.* 23, 468–475. doi: 10.3758/s13423-015-0896-2
- He, J., Guo, D., Zhai, S., Shen, M., and Gao, Z. (2019). Development of social working memory in preschoolers and its relation to theory of mind. *Child Dev.* 90, 1319–1332. doi: 10.1111/cdev.13025
- Jackson, M. C., and Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 556–568. doi: 10.1037/0096-1523.34.3.556
- Jiang, Y., Olson, I. R., and Chun, M. M. (2000). Organization of visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 683–702. doi: 10.1037/0278-7393.26.3.683
- Kaiser, D., Stein, T., and Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychon. Bull. Rev.* 22, 1784–1790. doi: 10.3758/s13423-015-0833-4
- Le Grand, R., Mondloch, C. J., Maurer, D., and Brent, H. P. (2001). Early visual experience and face processing. *Nature* 410, 890–890. doi: 10.1038/35073749
- Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006
- Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356. doi: 10.1038/nn.3655
- Xie, W., and Zhang, W. (2017). Discrete item-based and continuous configural representations in visual short-term memory. *Vis. Cogn.* 25, 21–33. doi: 10.1080/13506285.2017.1339157

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Magen, Peelen, Emmanouil and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Relation Between Working Memory Capacity of Biological Movements and Fluid Intelligence

Tian Ye<sup>1</sup>, Peng Li<sup>2</sup>, Qiong Zhang<sup>1</sup>, Quan Gu<sup>1</sup>, Xiqian Lu<sup>1</sup>, Zaifeng Gao<sup>1\*</sup> and Mowei Shen<sup>1\*</sup>

<sup>1</sup> Department of Psychology, Zhejiang University, Hangzhou, China, <sup>2</sup> School of Education and Management, Yunnan Normal University, Kunming, China

## OPEN ACCESS

### Edited by:

Pietro Spataro,  
Mercatorum University, Italy

### Reviewed by:

Alejandro Galvez-Pol,  
University College London,  
United Kingdom  
Caterina Artuso,  
University of Urbino Carlo Bo, Italy

### \*Correspondence:

Zaifeng Gao  
zaifengg@zju.edu.cn  
Mowei Shen  
mwshen@zju.edu.cn

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 01 August 2019

**Accepted:** 27 September 2019

**Published:** 18 October 2019

### Citation:

Ye T, Li P, Zhang Q, Gu Q, Lu X,  
Gao Z and Shen M (2019) Relation  
Between Working Memory Capacity  
of Biological Movements and Fluid  
Intelligence. *Front. Psychol.* 10:2313.  
doi: 10.3389/fpsyg.2019.02313

Studies have revealed that there is an independent buffer for holding biological movements (BM) in working memory (WM), and this BM-WM has a unique link to our social ability. However, it remains unknown as to whether the BM-WM also correlates to our cognitive abilities, such as fluid intelligence (Gf). Since BM processing has been considered as a hallmark of social cognition, which distinguishes from canonical cognitive abilities in many ways, it has been hypothesized that only canonical object-WM (e.g., memorizing color patches), but not BM-WM, emerges to have an intimate relation with Gf. We tested this prediction by measuring the relationship between WM capacity of BM and Gf. With two Gf measurements, we consistently found moderate correlations between BM-WM capacity, the score of both Raven's advanced progressive matrix (RAPM), and the Cattell culture fair intelligence test (CCFIT). This result revealed, for the first time, a close relation between WM and Gf with a social stimulus, and challenged the double-dissociation hypothesis for distinct functions of different WM buffers.

**Keywords:** biological motion, working memory, fluid intelligence, PLDs, IQ

## INTRODUCTION

Biological movements (BMs) refer to the movements of animate entities (Johansson, 1973). Researchers have demonstrated converging evidence that BM contains abundant social information; for example, identity, gender, social interaction, intention, and emotion can be extracted from BM (e.g., Pollick et al., 2002; Atkinson et al., 2004; Petrini et al., 2014; for reviews see Puce and Perrett, 2003; Blake and Shiffrar, 2007; Troje, 2013). The ability to successfully and efficiently process human BM is critical to being a functioning member of human society (e.g., Perry et al., 2010; Herrington et al., 2011; Pavlova, 2012; Troje, 2013; Cook et al., 2014; Ding et al., 2017; Thornton, 2018), and healthy adults are considered experts at processing BM (Johansson, 1973; Fox and McDaniel, 1982; Troje, 2013).

Our cognitive system even involves an independent buffer for processing BMs (Smyth et al., 1988; Smyth and Pendleton, 1989, 1990; Wood, 2007, 2008, 2011; Cortese and Rossi-Arnaud, 2010; Shen et al., 2014; Liu et al., 2019) in working memory (WM), which maintains and manipulates a limited amount of information for the ongoing tasks (Baddeley and Hitch, 1974; Cowan, 2010).

Just as there is a WM module specific for object or spatial information (i.e., object-WM vs. spatial-WM; Baddeley, 1996), there is also a WM module dedicated to BM information (Smyth et al., 1988; Smyth and Pendleton, 1989; Wood, 2007, 2008, 2011; Shen et al., 2014; Liu et al., 2019). Previous studies explored the BM-WM mechanisms by using real human movements (Smyth et al., 1988; Smyth and Pendleton, 1990; Wu and Coulson, 2014), computer-generated animations of human movements (Wood, 2007, 2011), imaginary BMs by the given names (Cortese and Rossi-Arnaud, 2010), and point light displays (PLDs) of human movements (Shen et al., 2014; Liu et al., 2019). For instance, Wood (2007) and Shen et al. (2014) demonstrated that participants could simultaneously hold a set of BM and a set of visual objects (e.g., colors) or locations in WM without significant mutual impairments. Liu et al. (2019) further found that memorizing BM was not modulated by the number of concurrent retained feature bindings in WM, and vice versa. Meanwhile, as compared to object-WM, BM-WM only holds a maximum of 3–4 BM stimuli (Smyth et al., 1988; Wood, 2007; Shen et al., 2014).<sup>1</sup> Later neuroimaging studies further uncovered the neural substrates of BM-WM by showing that the mirror neuron system (MNS) plays a pivotal role in retaining BM in WM (Gao et al., 2013; Lu et al., 2016; Cai et al., 2018). Recent studies have also begun to explore issues such as the development of BM-WM (He et al., 2019), the influence of other social information (e.g., social interaction and emotion) on BM-WM capacity (Ding et al., 2017; Guo et al., 2019), BM-related binding in WM (Wood, 2008; Poom, 2012; Ding et al., 2015; Gu et al., 2019), the representation format of BM in WM (Vicary and Stevens, 2014; Vicary et al., 2014), and the frame of reference for remembering BM (Wood, 2010).

Although working memory<sup>2</sup> capacity is rather limited, ample studies have consistently revealed that WM capacity has substantial predictive power in terms of predicting performance of high-level cognitive activities, including reading abilities, scholastic aptitude, information selection, and fluid intelligence (Gf) (e.g., Conway et al., 2002; Woodman et al., 2007; Hollingworth et al., 2008; Unsworth et al., 2014). Among these intimate relations, the relation between WM and Gf has received particular attention. Gf refers to the abilities needed for abstract reasoning and speeded performance (Cattell, 1971). In the last 15 years, researchers have revealed that the WM capacity of visual objects (e.g., color, shape; the corresponding WM buffer is named as object-WM) can significantly predict an individual's Gf (Kane et al., 2005; Fukuda et al., 2010; Unsworth et al., 2014; Hicks et al., 2015). However, no study thus far has investigated the relation between BM-WM capacity and Gf, the answer to which will significantly improve our understanding about the processing mechanism and the function of BM-WM.

<sup>1</sup>It is of note that there are two distinct views as to the unit of VWM (i.e., slots vs. resources; Bays and Husain, 2008; Zhang and Luck, 2008; see Suchow et al., 2014 for a review). So far, the available studies implicitly assumed that slots are basic units in VWM for storing BMs. Here we reviewed and summarized these results. The unit of storing BMs in VWM is beyond the scope of current work.

<sup>2</sup>The term WM used here is identical to the system that is often called short-term memory. Considering most of the extant studies using change detection task to explore the mechanisms of short-term memory adopted the term WM, we followed the previous studies and used the term WM.

On the one hand, there might be no relation between WM capacity of BM and Gf. It has been claimed that the recognition ability to process BM is a hallmark of social cognition (Pavlova, 2012). Neuroimaging studies have revealed that the human MNS, which serves as key neural substrates for social activities such as mentalizing and empathy (e.g., Gallese and Goldman, 1998; Hooker et al., 2008; Shamay-Tsoory et al., 2009; Liakakis et al., 2011; Grecucci et al., 2013; Sperduti et al., 2014), plays a pivotal role not only in visual perception of BM (Saygin et al., 2004; Pavlova, 2012; Gilaie-Dotan et al., 2013), but also in retaining BM in WM (Gao et al., 2013; Lu et al., 2016). Moreover, recent WM studies implied that even merely retaining a frame of BM (e.g., static hand gestures) in WM, the MNS is also involved (Galvez-Pol et al., 2018a,b; Arslanova et al., 2019). Since the BM-WM buffer is suggested to play an important role in transferring ongoing social information from perception to WM (Urgolites and Wood, 2013; Shen et al., 2014), it is possible that the BM-WM capacity could inherently predict one's social ability instead of the general cognitive ability (e.g., Gf). Corroborating this possibility, we recently found that BM-WM capacity positively correlated with both empathy (Gao et al., 2016) and theory of mind score (He et al., 2019), whereas such a relation vanished for WM capacity of movements of rectangles (i.e., non-animate motion) or of colors (i.e., object-WM). Because of the intimate relation between BM and social ability in both perception and WM, it has been suggested that BM-WM is a representative of social WM (He et al., 2019), which maintains and manipulates a limited set of social information in an online manner and is of paramount importance for navigating our social environment (Meyer and Lieberman, 2012), and is the best manner to measure the development of social WM in preschoolers (He et al., 2019). Critically, previous studies have only addressed whether there was a link between BM-WM and social ability, but no study has examined whether BM-WM is constrained to social ability. In other words, whether BM-WM capacity has no predictive power over Gf needs to be elucidated. If a null result is revealed, we then find a double dissociation in terms of different roles of WM buffers, with object-WM closely linking to canonical cognitive ability and BM-WM correlating to social ability.

On the other hand, since the storage of WM involves a series of cognitive operations, WM may have a tight relation with Gf, regardless of the stimuli type. Two recent functional magnetic resonance (fMRI) studies (Lu et al., 2016; Cai et al., 2018) found that, in addition to the MNS, the superior and inferior parietal lobule (SPL and IPL) and bilateral prefrontal cortex, which contribute to general cognitive processes (e.g., Todd and Marois, 2004; Xu and Chun, 2006; Barbey et al., 2013), also play a role in retaining BM in WM. Therefore, it is also possible that BM-WM capacity not only correlates to social ability, but also links to general cognitive ability.

The current study thus attempted to elucidate whether BM-WM has a close relationship with Gf. BM-WM was measured by using PLDs stimuli (Johansson, 1973). To ensure the validity of our study and facilitate comparisons with previous research, we adopted the widely used Raven advanced progressive matrices (RAPM) and Cattell cultural fair intelligence test (CCFIT) as



our Gf measurements (Kane et al., 2005; Fukuda et al., 2010; Unsworth et al., 2014).

## PILOT STUDY

We first conducted a pilot study with 60 participants to estimate the potential correlation coefficients between BM-WM and the Gf measurements. Results from the pilot study were then used for calculating the final sample size with 90% power on a 0.05 significant level.

## Method

### Participants

A total number of 60 participants took part in the pilot study. Thirty (18 males; mean  $\pm$  SD age  $21.3 \pm 2.04$  years) participants were from Zhejiang University, and thirty (15 males; mean  $\pm$  SD age  $18.9 \pm 1.03$  years) were from Yunnan University and Yunnan Normal University. Participants all had normal color vision and normal or corrected-to-normal visual acuity and received payment/course credit for their participation. Before the experiment, participants provided signed informed consent. The study was approved by the Research Ethics Board of Zhejiang University, Yunnan University and Yunnan Normal University.

### Stimuli and Apparatus

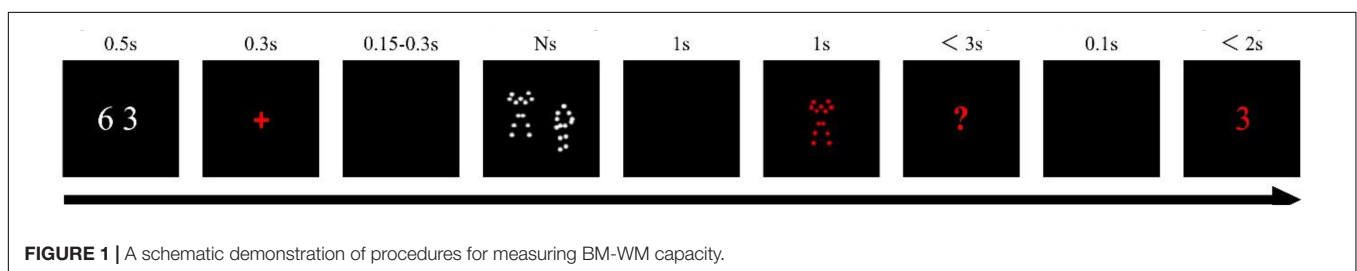
For the BM-WM test, PLDs were used as the BM stimuli. For each PLDs movement, 13 light points are placed at distinct joints of a moving human body to form a coherent and meaningful movement. We adopted PLDs in order to isolate BM information from other sources (e.g., color, contour, and texture; for a review see Troje, 2013). Seven movements were selected from the Vanrie and Verfaillie (2004) database: cycling, jumping, painting, spading, walking, waving, and chopping (see **Figure 1**). 30 distinct frames consisted one animation with each animation displayed twice consecutively, leading to a 1-s animation (refresh rate, 60 Hz). Each stimulus subtended a visual angle of approximately  $1.64^\circ \times 1.64^\circ$  from a viewing distance of 60 cm. In line with previous studies measuring BM capacity (e.g., Shen et al., 2014; Gao et al., 2016), during each trial one to five distinct stimuli would show up randomly on the periphery of an invisible circle (radius,  $4.88^\circ$  from the screen center) evenly.

For Gf measurement, two solidly validated Gf questionnaires were adopted: the Cattell culture fair intelligence test (CCFIT) and Raven's advanced progressive matrix (RAPM). These two questionnaires were chosen for two considerations. First, both are non-verbal tests, which enable us to largely remove the

influence from different culture-backgrounds. Second, both have been widely used in measuring the relation between WM and Gf (e.g., Kane et al., 2005; Fukuda et al., 2010; Unsworth et al., 2014). For CCFIT, we adopted a full-scale measure in accordance with previous studies (Fukuda et al., 2010; Unsworth et al., 2014), which is composed of four separate and timed paper-and-pencil sessions (Cattell, 1971). Participants were given about 2~3.5 min to finish each session. In the first session, participants saw 13 incomplete series of abstract shapes, along with 6 alternatives for each, and selected one that best completed the series. In the second session, participants saw 14 problems composed of abstract shapes, and chose the two out of the five that differed from the other three, e.g., shapes differed in content, orientation, or size. In the third session, participants saw 13 incomplete matrices containing four to nine boxes that had abstract shapes as well as an empty box and six choices. They had to infer the relations among the items in the matrix and select an answer that correctly fulfill each matrix. In the fourth session, participants saw 10 sets of abstract figures consisting of lines and a single dot along with five alternatives. They needed to assess the relation among the dot, figures, and lines, and choose the alternative in which a dot could be placed according to the same relation. The final score of CCFIT was the total number of items solved correctly across all four sessions. For RAPM, which is a measure of abstract reasoning, we chose a split-half measure (Jaeggi et al., 2008; Broadway and Engle, 2010; Shipstead et al., 2012) to shorten total experiment time course to avoid fatigue. A full scale of RAPM was split into odds and even items and each participant had 20 min to complete the split-half scale. Note that in previous studies wherein a split-half measure of RAPM was adopted, researchers usually gave participants 30 min to finish the test. However, a pre-test with a sample of another 10 participants from Zhejiang University resulted in ceiling effect with a 30-min duration, we hence reduced the testing duration to 20 min. Each split-half measure of RAMP consists of 18 items displayed in ascending order of difficulty. Each item consists of a display of  $3 \times 3$  matrices of geometric patterns with the bottom right one missing. Participant had to select one among eight alternatives, which can correctly complete the overall series of patterns. The final score of RAPM was the total number of correct solutions. For both CCFIT and RAPM, participants scored 1 point if they answered correctly on an item and 0 if they were wrong.

### Procedure

For the BM-WM test, each trial began with two white digits showing in the center of the screen for 500 ms (see **Figure 1**).



Participants were demanded to repeat the two digits (e.g., “six,” “three,” “six,” and “three”) aloud. This manipulation was set to prevent them from verbally processing those movements (cf. Gao et al., 2013; Shen et al., 2014; van Boxtel and Lu, 2015). A red fixation then appeared for 300 ms and, after a blank interval of 150 ms–350 ms, the memory array was presented on the screen for *Ns* (according to the number of PLDs movements, e.g., 5 s for 5 stimuli; cf. Shen et al., 2014) to avoid underestimating the WM capacity of BM. Following a 1-s blank interval, a red probe appeared in the screen center for 1 s. From then on, participants stopped repeating the digits. The probe then disappeared, followed by a red question mark showing at the screen center, and participants had 3 s to decide whether the probe had appeared in the memory array before by pressing a button to relay the judgment. After the response, or if no response was made within 3 s, a red digit would be presented after a 100 ms delay. Participants had to decide whether the red digit was one of the previously rehearsed digits by pressing the same buttons as above. Participants were told to respond as accurately as possible. There were 30 trials under each memory load, resulting in 150 trials in total. Before the formal experiment, there are 16 trials for participants to practice.

Half of the participants performed the BM-WM measurement before the two *Gf* measurements and the other half on the opposite, and the two *Gf* measurements were given to the participants in random order. Before each task, the experimenter would stress to the participant that they needed to try their best, either to remember the stimuli or to answer each item in the two questionnaires. For *Gf* measurement, participants were instructed to write their answers on an answering sheet and draft papers were provided. The experimenter monitored the time to ensure the task was fulfilled within the required time window. The whole test was around 70 min.

## Analysis

To estimate BM-WM capacity, we employed Cowan's formula (Cowan, 2001):  $K = S \times (H - F)$ , where *K* is the WM capacity, *S* is the number of to-be-memorized stimuli, *H* is the hit rate that refers to the successful detection of a new stimulus, and *F* is the false alarm rate that refers to an incorrect new-stimulus response. We calculated *K* for each set size of each participant. To have a more accurate estimate, we considered the maximum *K* (*K*-max) among the five load conditions as one's WM capacity (e.g., Curby and Gauthier, 2007; Gao et al., 2013; Shen et al., 2014; Guo et al., 2019). Only trials with correct responses for digit task were analyzed. Finally, Pearson's correlations between *K*-max and the scores on the two *Gf* measurements were calculated.

## Results

Descriptive statistics of each measured variable are shown in **Table 1**. One-sample Kolmogorov-Smirnov test showed that all the measured variables conformed to normal distributions ( $p$ s > 0.05; see also the Skewness in **Table 1**).

The overall accuracy of the digit rehearsal task was 95%. The correlation between CCFIT score and RAPM score was  $r = 0.597$ ,  $p < 0.001$ . Pearson's correlation test revealed a

**TABLE 1** | Mean value (SE) and results of skewness test of each measured variable in the current study.

	Statistics	WM task ( <i>K</i> -max)	CCFIT Score	RAPM Score
<b>Pilot study</b>				
BM	Mean (SE)	3.05 (0.73)	37.37 (5.16)	11.93 (2.67)
	Skewness	0.02	−0.76	−0.15
<b>Formal study</b>				
BM	Mean (SE)	3.14 (0.77)	37.40 (4.67)	11.61 (2.73)
	Skewness	−0.01	−0.63	−0.38

significantly positive correlation between *K*-max and CCFIT score (**Figure 2A**),  $r = 0.643$ ,  $p < 0.001$ , as well as between *K*-max and the RAPM score (**Figure 2B**),  $r = 0.594$ ,  $p < 0.001$ .

## Discussion

Results of our pilot study revealed a significant correlation between BM-WM capacity and *Gf*, suggesting that the performance of BM-WM can predict one's cognitive ability. As a small sample size of 60 may not be sufficient to draw a robust conclusion, we used G\*power 3.1 to determine our final sample size (Faul et al., 2009). To achieve a medium effect size ( $d = 0.3$  for Pearson correlation) and a power of 0.9 at 0.05 significant level, we had to test at least 112 participants. To this end, we tested another 55 participants to ensure our sample size is big enough. All testing procedures were pre-registered with the Open Science Framework<sup>3</sup>.

## FORMAL STUDY

### Method

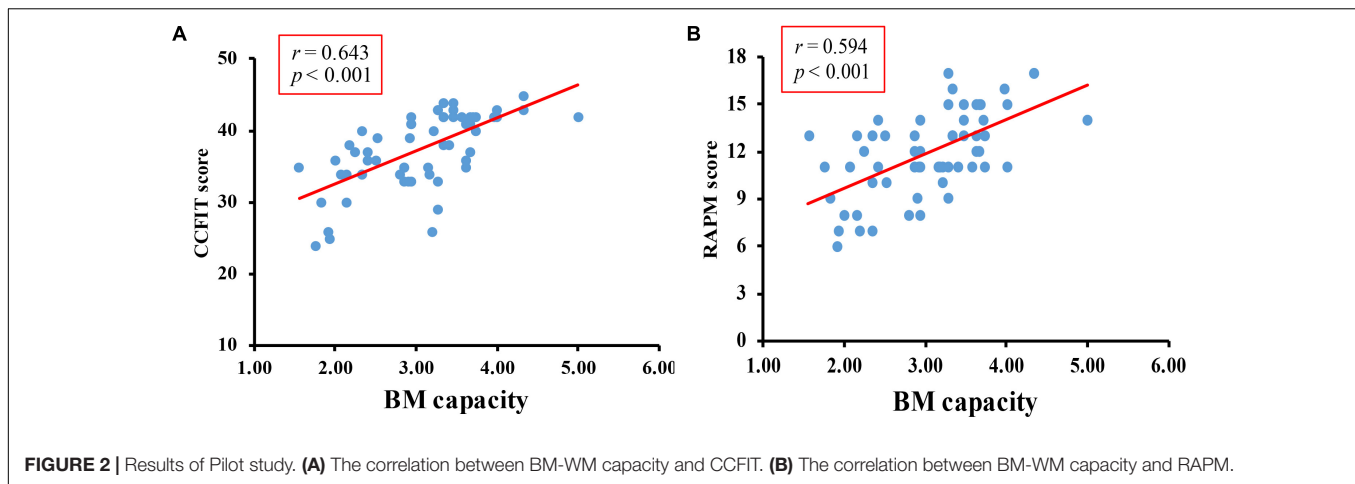
Together with the 60 participants in the pilot study, 115 (60 female; mean  $\pm$  SD age  $20.1 \pm 1.7$  years) participants took part in the experiment. Eight-five participants were from Zhejiang University and 30 were from Yunnan University and Yunnan Normal University. Participants all had normal or correct-to-normal vision and normal color vision. Participants received payment/course credit for their participation. Two participants were excluded from further analysis because the *K*-max was below 3 standard deviations of the average, which resulted in a final sample size of 113. Before the experiment, participants provided signed informed consent. The study was approved by the Research Ethics Board of Zhejiang University, Yunnan University, and Yunnan Normal University. The stimuli and procedures were all the same as in the pilot study.

### All Results

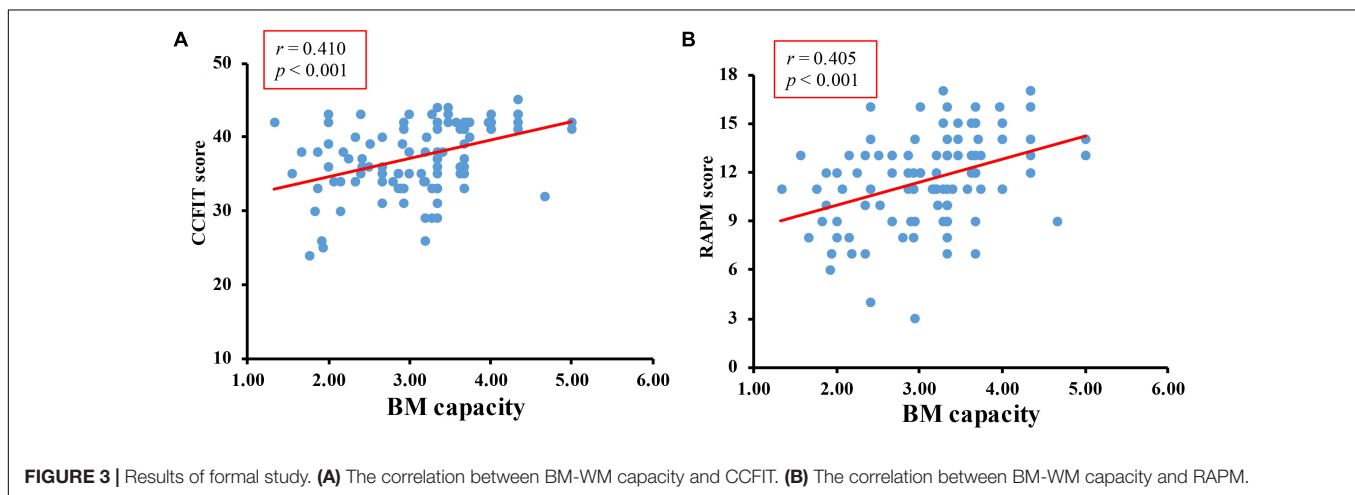
Descriptive statistics of each measured variable and results of tests for skewness are shown in **Table 1**.

Overall accuracy of the digit rehearsal task was 96%. The correlation between CCFIT score and RAPM score was  $r = 0.579$ ,  $p < 0.001$ . The correlations between *K*-max and CCFIT, *K*-max and RAPM were  $r = 0.410$ ,  $p < 0.001$ , and  $r = 0.405$ ,  $p < 0.001$ , respectively (**Figure 3**).

<sup>3</sup><https://osf.io/cqkx6/>



**FIGURE 2 |** Results of Pilot study. **(A)** The correlation between BM-WM capacity and CCFIT. **(B)** The correlation between BM-WM capacity and RAPM.



**FIGURE 3 |** Results of formal study. **(A)** The correlation between BM-WM capacity and CCFIT. **(B)** The correlation between BM-WM capacity and RAPM.

## GENERAL DISCUSSION

The goal of our study was to examine whether BM-WM capacity can predict canonical cognitive ability. In contrast to the prediction of a null relation between BM-WM and Gf, correlation analysis revealed a significantly positive correlation between BM-WM capacity and the two Gf measurements, suggesting that, although BM processing has an intimate relation with social cognition, the capacity of BM-WM can predict one's high-level general cognitive ability (Gf).

### Why a Relation Between BM-WM and Gf Exists?

We argue that the reason for an intimate relation between BM-WM and Gf may lie in the neural substrates involved in BM processing. While distinct visual objects (e.g., color and shape) are processed and stored via the primary visual cortex (Harrison and Tong, 2009; Serences et al., 2009), the processing of BM requires the involvement of a much broader brain network. Neuroimaging studies revealed that two visual pathways are engaged in BM processing, with the ventral pathway handling form information while the dorsal pathway

addressing motion information (e.g., Giese and Poggio, 2003; Gilaie-Dotan et al., 2011). The two pathways converge in the superior temporal sulcus (STS) to have a coherent representation of BM. Additionally, the MNS is also involved in processing BM (e.g., Rizzolatti and Craighero, 2004; Pineda, 2005; Perry et al., 2010). Along the same lines, recent fMRI studies revealed that both the neural substrates that were dedicated to core social ability (MNS), and those for canonical cognitive processing (SPL, IPL, and bilateral prefrontal cortex), are involved in the retention of BM in WM (Lu et al., 2016; Cai et al., 2018). Therefore, we consider that our previous work showing the relation between BW-WM and empathy reflects the contribution of MNS as well as STS, and the current finding may reflect the contribution of SPL, IPL, and bilateral prefrontal cortex. From this perspective, the current study sheds critical light on future clinical interventions focusing on WM. That is, future clinical interventions might consider training on BM-WM, which might be beneficial to both cognitive and social abilities.

### Implications of the Current Study

The current study is among the first that directly investigates the relationship between BM-WM and Gf, contributing to the



BM research field in general and the BM-WM explorations in particular. Although there have been a few studies examining the relationship between BM perception and Gf, the results were mixed (e.g., Barresi and Moore, 1996; Shinkfield et al., 1999; Blake et al., 2003; Rutherford and Troje, 2012). Recently researchers even considered Gf to play a “scaffolding” role in processing BM, i.e., when one’s social ability is impaired, individuals turn to exploit general cognitive processes to handle BM information (Rutherford and Troje, 2012). The current study extended the exploration from perception to WM. In contrast to the implications from BM perception, we presented clear-cut evidence that higher WM capacity of BM predicts a higher IQ score. Therefore, the BM-WM capacity not only correlates to social ability but also has an intimate link with cognitive ability.

The current exploration also shed critical light on the function of social WM. Currently, it has been revealed that human brain has evolved neural substrates dedicated to social WM (e.g., dorsomedial prefrontal cortex, ventromedial prefrontal cortex, and right temporo-parietal junction; Meyer and Lieberman, 2012; Meyer et al., 2012, 2015), which is deactivated during canonical cognitive WM tasks (e.g., memorizing colors, locations, letters). Although previous social WM studies focused on peoples’ trait and emotions (e.g., Meyer and Lieberman, 2012, 2016; Meyer et al., 2012, 2015; Thornton and Conway, 2013; Xin and Lei, 2015), the explorations of social WM should not be constrained to these sets of information. Indeed, the advance of social WM is to emphasize that the canonical WM studies have largely overlooked the temporal storage and manipulation of social information in our life, for instance, people’s identities, mental states, traits, and interpersonal relationships. As we reviewed in the introduction, a bunch of social information (identity, emotion, intention, goal, and gender, etc.) could be extracted even from PLD format BM, and one’s recognition ability of BM is taken as a hallmark of social cognition (Pavlova, 2012), which, to the best of our knowledge, is the only stimuli category receiving this evaluation in terms of measuring social ability. To this end, we consider that BM is a well-matched stimulus in measuring social WM, and we used it to measure the development of social WM in 3~6 preschoolers (He et al., 2019). Taking all the related explorations of social WM together (i.e., using people’s trait, emotion, and BM as the stimuli of interest), we noticed that the extant studies on social WM mainly focused on the storage capacity and manner (Shen et al., 2014; Gao et al., 2016; He et al., 2019; Liu et al., 2019), and neural substrates of social WM (Lieberman, 2007; Meyer and Lieberman, 2012, 2016; Meyer et al., 2012, 2015; Thornton and Conway, 2013; Xin and Lei, 2015; Lu et al., 2016). A few studies had attempted to explore the functions of social WM (Meyer et al., 2012, 2015; Xin and Lei, 2015; Gao et al., 2016). However, to date, all of related studies focused on the relation of social WM to social abilities. The current study closed a key gap when uncovering the function of social WM, and implied that, although social WM had special neural substrates from canonical cognitive WM (e.g., object WM), there were no double dissociations in terms of different roles of WM buffers (i.e., canonical cognitive WM links to cognitive ability and social WM links to social

ability). Instead, akin to canonical cognitive WM, the capacity of social WM (at least for certain representatives) has a close relationship with Gf.

## Limitations & Future Studies

The current study aimed at exploring the function of BM-WM by exploring the relationship between BM-WM capacity and Gf using a correlation analysis. To have a comprehensive understanding of the function of BM-WM, we argue that at least two aspects have to be addressed in future studies. First, additional study is needed to further examine the relation between BM-WM capacity and Gf, for instance, by using different testing procedures (e.g., a recall task of WM, Zhang and Luck, 2008) and sample selections (e.g., using students in primary or middle school). Moreover, the current experiment essentially used a dual-task setting (i.e., an articulatory suppression task and a WM task), which has been widely used in both BM-WM and object-WM fields to measure the WM capacity. Future study may consider to partial out the effect of articulatory suppression, such that we can have a pure estimation of the relation between BM-WM capacity and Gf. Second, Gao et al. (2016) and the current study explored the function of BM-WM from the perspective of social ability and cognitive ability, respectively. Moreover, both studies used a correlational analysis. To have a full picture of the function of BM-WM and to figure out the underlying relation between social and cognitive abilities, it would be more informative to put all the related factors (e.g., empathy, Gf, BM-WM capacity, object-WM capacity, and attention) in one study, and perform more comprehensive analysis such as latent variable analysis (e.g., Unsworth et al., 2014).

Additionally, based on the previous studies (Smyth et al., 1988; Smyth and Pendleton, 1989, 1990; Wood, 2007, 2008, 2011; Cortese and Rossi-Arnaud, 2010; Shen et al., 2014; Liu et al., 2019), the current study claimed that BM has an independent buffer in WM in terms of Baddeley and Hitch (1974). It is worth noting that processing (including perception and WM) human body-related images (e.g., hand gesture) also activates MNS (Mecklinger et al., 2002; Moreau, 2013; Galvez-Pol et al., 2018a,b; Arslanova et al., 2019), hence, it is also possible that the currently tapped BM-WM is actually an independent buffer dedicated to maintaining body-related stimuli regardless of motion (BM is just one example). However, we argue that it is premature to claim an independent WM buffer for body-related stimuli, considering that all related studies on the storage buffer of BM in WM focused on dynamic BM (Smyth et al., 1988; Smyth and Pendleton, 1989, 1990; Wood, 2007, 2008, 2011; Cortese and Rossi-Arnaud, 2010; Shen et al., 2014; Liu et al., 2019). Indeed, there are at least two reasons against the use of this independent WM buffer for body-related stimuli in general. First, the processing of BM is more complex than a single body-related image in terms of both cognitive and neural processing. For cognitive processing, the formation of a coherent BM representation requires our cognitive system to integrate different pieces of information (i.e., individual frames or images) across space and time (Orgs and Haggard, 2011). Lange and Lappe (2006) suggested that BM

perception is achieved by dynamically integrating the activity of template cells of static form information the human body (i.e., body image), and this process requires the help of attention (Thornton et al., 2002; see Thompson and Parasuraman, 2012 for a review). For neural processing, unlike perceiving hand or face images which usually activates more posterior cortices, such as somatosensory cortices, extrastriate body area, and fusiform (Kanwisher et al., 1997; McCarthy et al., 1997; Gauthier et al., 2000; Galvez-Pol et al., 2018a,b; Arslanova et al., 2019), BM perception and WM maintenance activate more anterior regions, such as superior temporal sulcus, inferior frontal gyrus and ventral premotor cortex (e.g., Perry et al., 2010; Lu et al., 2016; Cai et al., 2018). Second, according to the core knowledge architecture of visual WM of Wood (2011), BM and body-related image should be stored in different buffers. This architecture claims that there are distinct buffers in visual WM for retaining spatiotemporal information (for object tracking, e.g., BM), object property/kind information (for object recognition, e.g., the form of a BM stimulus), and view-dependent snapshot information (for place recognition; e.g., four distinct colors in a 2D space). The dynamic BM belongs to spatiotemporal information, while body-related image belongs to view-dependent snapshot information. To this end, we consider that it is important to examine whether there is an independent WM buffer for body-related stimuli, by requiring participants to memorize dynamic BM and static body-related stimuli in one task.

## DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/supplementary files.

## REFERENCES

- Arslanova, I., Galvez-Pol, A., Calvo-Merino, B., and Forster, B. (2019). Searching for bodies: ERP evidence for independent somatosensory processing during visual search for body-related information. *Neuroimage* 195, 140–149. doi: 10.1016/j.neuroimage.2019.03.037
- Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., and Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33, 717–746. doi: 10.1068/p5096
- Baddeley, A. (1996). The fractionation of working memory. *Proc. Natl. Acad. Sci.* 93, 13468–13472.
- Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89.
- Barbey, A. K., Koenigs, M. R., and Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex* 49, 1195–1205. doi: 10.1016/j.cortex.2012.05.022
- Barresi, J., and Moore, C. (1996). Intentional relations and social understanding. *Behav. Brain Sci.* 19, 107–122. doi: 10.1017/s0140525x00041790
- Bays, P. M., and Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854. doi: 10.1126/science.1158023
- Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Ann. Rev. Psychol.* 58, 47–73.
- Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., and Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychol. Science* 14, 151–157. doi: 10.1111/1467-9280.01434
- Broadway, J. M., and Engle, R. W. (2010). Validating running memory span: measurement of working memory capacity and links with fluid intelligence. *Behav. Res. Methods* 42, 563–570. doi: 10.3758/BRM.42.2.563
- Cai, Y., Urgolites, Z., Wood, J., Chen, C., Li, S., Chen, A., et al. (2018). Distinct neural substrates for visual short-term memory of actions. *Hum. Brain Mapp.* 39, 4119–4133. doi: 10.1002/hbm.24236
- Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action*. Boston: Houghton Mifflin.
- Conway, A. R., Cowan, N., Bunting, M. F., Theriault, D. J., and Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence* 30, 163–183. doi: 10.1016/s0160-2896(01)00096-4
- Cook, J., Swapp, D., Pan, X., Bianchi-Berthouze, N., and Blakemore, S. J. (2014). Atypical interference effect of action observation in autism spectrum conditions. *Psychol. Med.* 44, 731–740. doi: 10.1017/S0033291713001335
- Cortese, A., and Rossi-Arnaud, C. (2010). Working memory for ballet moves and spatial locations in professional ballet dancers. *Appl. Cogn. Psychol.* 24, 266–286. doi: 10.1002/acp.1593
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. doi: 10.1017/S0140525X01003922
- Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* 19, 51–57. doi: 10.1177/0963721409359277
- Curby, K. M., and Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychon. Bull. Rev.* 14, 620–628. doi: 10.3758/bf03196811

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Board of Zhejiang University, Yunnan University, and Yunnan Normal University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

ZG put up the question and designed the experiment. TY and PL collected the data. QZ helped with preparing IQ measurements. TY wrote the first draft of the manuscript. ZG made critical changes on it. QG, XL, and MS provided the meaningful suggestions on the final version of the manuscript.

## FUNDING

This research was supported by the National Natural Science Foundation of China grants 31771202 and 31571119, the Project of Ministry of Science and Technology of the People's Republic of China (2016YFE0130400), Medical and Health Science Research Fund of Zhejiang Province (Nos. 2017KY352 and 2018KY064), and the MOE Project of Humanities and Social Sciences (No. 17YJA190005).

## ACKNOWLEDGMENTS

We appreciate the help from Xiaochi Ma and Xiaoyuan Yang in data collection.

- Ding, X., Gao, Z., and Shen, M. (2017). Two equals one: two human actions during social interaction are grouped as one unit in working memory. *Psychol. Sci.* 28, 1311–1320. doi: 10.1177/0956797617707318
- Ding, X., Zhao, Y., Wu, F., Lu, X., Gao, Z., and Shen, M. (2015). Binding biological motion and visual features in working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 850–865. doi: 10.1037/xhp0000061
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A. G. (2009). Statistical power analyses using G\* Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/brm.41.4.1149
- Fox, R., and McDaniel, C. (1982). The perception of biological motion by human infants. *Science* 218, 486–487. doi: 10.1126/science.7123249
- Fukuda, K., Vogel, E., Mayr, U., and Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychon. Bull. Rev.* 17, 673–679. doi: 10.3758/17.5.673
- Gallese, V., and Goldman, A. (1998). Mirror neurons and the mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/s1364-6613(98)01262-5
- Galvez-Pol, A., Calvo-Merino, B., Capilla, A., and Forster, B. (2018a). Persistent recruitment of somatosensory cortex during active maintenance of hand images in working memory. *Neuroimage* 174, 153–163. doi: 10.1016/j.neuroimage.2018.03.024
- Galvez-Pol, A., Forster, B., and Calvo-Merino, B. (2018b). Modulation of motor cortex activity in a visual working memory task of hand images. *Neuropsychologia* 117, 75–83. doi: 10.1016/j.neuropsychologia.2018.05.005
- Gao, Z., Bentin, S., and Shen, M. (2013). Rehearsing biological motion in working memory: an EEG study. *Psychologist* 26, 194–198. doi: 10.1162/jocn\_a\_00687
- Gao, Z., Ye, T., Shen, M., and Perry, A. (2016). Working memory capacity of biological movements predicts empathy traits. *Psychon. Bull. Rev.* 23, 468–475. doi: 10.3758/s13423-015-0896-2
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *J. Cogn. Neurosci.* 12, 495–504. doi: 10.1162/089892900562165
- Giese, M. A., and Poggio, T. (2003). Cognitive neuroscience: neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* 4, 179–192. doi: 10.1016/j.bbi.2015.01.001
- Gilaie-Dotan, S., Bentin, S., Harel, M., Rees, G., and Saygin, A. P. (2011). Normal form from biological motion despite impaired ventral stream function. *Neuropsychologia* 49, 1033–1043. doi: 10.1016/j.neuropsychologia.2011.01.009
- Gilaie-Dotan, S., Kanai, R., Bahrami, B., Rees, G., and Saygin, A. P. (2013). Neuroanatomical correlates of biological motion detection. *Neuropsychologia* 51, 457–463. doi: 10.1016/j.neuropsychologia.2012.11.027
- Grecucci, A., Giorgetta, C., Bonini, N., and Sanfey, A. G. (2013). Reappraising social emotions: the role of inferior frontal gyrus, temporo-parietal junction and insula in interpersonal emotion regulation. *Front. Hum. Neurosci.* 7:523. doi: 10.3389/fnhum.2013.00523
- Gu, Q., Li, W., Lu, X., Chen, H., Shen, M., and Gao, Z. (2019). Agent identity drives adaptive encoding of biological motion into working memory. *J. Vis.* 18:703. doi: 10.1167/18.703
- Guo, Y., Li, W., Lu, X., Xu, X., Qiu, F., Shen, M., et al. (2019). Emotional states affect the retention of biological motion in working memory. *Emotion* doi: 10.1037/emo0000668 [Epub ahead of print].
- Harrison, S. A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635. doi: 10.1038/nature07832
- He, J., Guo, D., Zhai, S., Shen, M., and Gao, Z. (2019). Development of social working memory in preschoolers and its relation to theory of mind. *Child Dev.* 90, 1319–1332. doi: 10.1111/cdev.13025
- Herrington, J. D., Nymberg, C., and Schultz, R. T. (2011). Biological motion task performance predicts superior temporal sulcus activity. *Brain Cogn.* 77, 372–381. doi: 10.1016/j.bandc.2011.09.001
- Hicks, K. L., Harrison, T. L., and Engle, R. W. (2015). Wonderlic, working memory capacity, and fluid intelligence. *Intelligence* 50, 186–195. doi: 10.1016/j.intell.2015.03.005
- Hollingworth, A., Richard, A. M., and Luck, S. J. (2008). Understanding the function of visual short-term memory: transsaccadic memory, object correspondence, and gaze correction. *J. Exp. Psychol.* 137, 163–181. doi: 10.1037/0096-3445.137.1.163
- Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., and D’Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Soc. Cogn. Affect. Neurosci.* 3, 204–217. doi: 10.1093/scan/nsn019
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6829–6833. doi: 10.1073/pnas.0801268105
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211. doi: 10.3758/bf03212378
- Kane, M. J., Hambrick, D. Z., and Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychol. Bull.* 131, 66–71. doi: 10.1037/0033-2909.131.1.66
- Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311. doi: 10.1523/jneurosci.17-11-04302.1997
- Lange, J., and Lappe, M. (2006). A model of biological motion perception from configural form cues. *J. Neurosci.* 26, 2894–2906. doi: 10.1523/jneurosci.4915-05.2006
- Liakakis, G., Nickel, J., and Seitz, R. J. (2011). Diversity of the inferior frontal gyrus-A meta-analysis of neuroimaging studies. *Behav. Brain Res.* 225, 341–347. doi: 10.1016/j.bbr.2011.06.022
- Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* 58, 259–289. doi: 10.1146/annurev.psych.58.110405.085654
- Liu, Y., Lu, X., Wu, F., Shen, M., and Gao, Z. (2019). Biological motion is stored independently from bound representation in working memory. *Vis. Cogn.* 1–13. doi: 10.1080/13506285.2019.1638479
- Lu, X., Huang, J., Yi, Y., Shen, M., Weng, X., and Gao, Z. (2016). Holding biological motion in working memory: an fMRI study. *Front. Hum. Neurosci.* 10:251. doi: 10.3389/fnhum.2016.00251
- McCarthy, G., Puce, A., Gore, J. C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 9, 605–610. doi: 10.1162/jocn.1997.9.5.605
- Mecklinger, A., Gruenewald, C., Besson, M., Magnie, M. N., and Von Cramon, D. Y. (2002). Separable neuronal circuitries for manipulable and non-manipulable objects in working memory. *Cereb. Cortex* 12, 1115–1123. doi: 10.1093/cercor/12.11.1115
- Meyer, M. L., and Lieberman, M. D. (2012). Social working memory: neurocognitive networks and directions for future research. *Front. Psychol.* 3:571. doi: 10.3389/fpsyg.2012.00571
- Meyer, M. L., and Lieberman, M. D. (2016). Social working memory training improves perspective-taking accuracy. *Soc. Psychol. Pers. Sci.* 7, 381–389. doi: 10.1177/194850615624143
- Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., and Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proc. Natl. Acad. Sci.* 109, 1883–1888. doi: 10.1073/pnas.1121077109
- Meyer, M. L., Taylor, S. E., and Lieberman, M. D. (2015). Social working memory and its distinctive link to social cognitive ability: an fMRI study. *Soc. Cogn. Affect. Neurosci.* 10, 1338–1347. doi: 10.1093/scan/nsv065
- Moreau, D. (2013). Motor expertise modulates movement processing in working memory. *Acta Psychol.* 142, 356–361. doi: 10.1016/j.actpsy.2013.01.011
- Orgs, G., and Haggard, P. (2011). Temporal binding during apparent movement of the human body. *Vis. Cogn.* 19, 833–845. doi: 10.1080/13506285.2011.598481
- Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cereb. Cortex* 22, 981–995. doi: 10.1093/cercor/bhr156
- Perry, A., Troje, N. F., and Bentin, S. (2010). Exploring motor system contributions to the perception of social information: evidence from EEG activity in the mu/alpha frequency range. *Soc. Neurosci.* 5, 272–284. doi: 10.1080/17470910903395767
- Petrini, K., Piwek, L., Crabbe, F., Pollick, F. E., and Garrod, S. (2014). Look at those two!: the precuneus role in unattended third-person perspective of social interactions. *Hum. Brain Mapp.* 35, 5190–5203. doi: 10.1002/hbm.22543
- Pineda, J. A. (2005). The functional significance of mu rhythms: translating “seeing” and “hearing” into “doing”. *Brain Res. Rev.* 50, 57–68. doi: 10.1016/j.brainresrev.2005.04.005
- Pollick, F. E., Lestou, V., Ryu, J., and Cho, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vis. Res.* 42, 2345–2355. doi: 10.1016/s0042-6989(02)00196-7



- Poom, L. (2012). Memory of gender and gait direction from biological motion: gender fades away but directions stay. *J. Exp. Psychol.* 38:1091. doi: 10.1037/a0028503
- Puce, A., and Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philos. Trans. R. Soc. Lond. Ser. B* 358, 435–445. doi: 10.1098/rstb.2002.1221
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.
- Rutherford, M. D., and Troje, N. F. (2012). IQ predicts biological motion perception in autism spectrum disorders. *J. Autism Dev. Disord.* 42, 557–565. doi: 10.1007/s10803-011-1267-0
- Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E., and Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *J. Neurosci.* 24, 6181–6188. doi: 10.1523/jneurosci.0504-04.2004
- Serences, J., Ester, E. E. F., Vogel, E. E. K., and Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 20, 207–214. doi: 10.1111/j.1467-9280.2009.02276.x
- Shamay-Tsoory, S. G., Aharon-Peretz, J., and Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain* 132, 617–627. doi: 10.1093/brain/awn279
- Shen, M., Gao, Z., Ding, X., Zhou, B., and Huang, X. (2014). Holding biological motion information in working holding biological motion information in working memory. *J. Exp. Psychol.* 40, 1332–1345. doi: 10.1037/a0036839
- Shinkfield, A. J., Day, R. H., Sparrow, W. A., and Zeman, L. (1999). Visual perception of human activity and gender in biological-motion displays by individuals with mental retardation. *Am. J. Ment. Retard.* 104, 215–226.
- Shipstead, Z., Redick, T. S., and Engle, R. W. (2012). Is working memory training effective? *Psychol. Bull.* 138, 628–654. doi: 10.1037/a0027473
- Smyth, M. M., Pearson, N. A., and Pendleton, L. R. (1988). Movement and working memory: patterns and positions in space. *Q. J. Exp. Psychol. A* 40, 497–514. doi: 10.1080/02724988843000041
- Smyth, M. M., and Pendleton, L. R. (1989). Working memory for movements. *Q. J. Exp. Psychol. A* 41, 235–250.
- Smyth, M. M., and Pendleton, L. R. (1990). Space and movement in working memory. *Q. J. Exp. Psychol. A* 42, 291–304.
- Sperduti, M., Guionnet, S., Fossati, P., and Nadel, J. (2014). Mirror neuron system and mentalizing system connect during online social interaction. *Cogn. Process.* 15, 307–316. doi: 10.1007/s10339-014-0600-x
- Suchow, J. W., Fougny, D., Brady, T. F., and Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Atten. Percept. Psychophys.* 76, 2071–2079. doi: 10.3758/s13414-014-0690-7
- Thornton, I. M. (2018). Stepping into the genetics of biological motion processing. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1687–1689. doi: 10.1073/pnas.1722625115
- Thornton, I. M., Rensink, R. A., and Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception* 31, 837–853. doi: 10.1068/p3072
- Thompson, J., and Parasuraman, R. (2012). Attention, biological motion, and action recognition. *Neuroimage* 59, 4–13. doi: 10.1016/j.neuroimage.2011.05.044
- Thornton, M. A., and Conway, A. R. (2013). Working memory for social information: chunking or domain-specific buffer? *Neuroimage* 70, 233–239. doi: 10.1016/j.neuroimage.2012.12.063
- Todd, J. J., and Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751–754. doi: 10.1038/nature02466
- Troje, N. F. (2013). “What is biological motion? Definition, stimuli and paradigms,” in *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*, eds M. D. Rutherford, and V. A. Kuhlmeier, (Cambridge, MA: MIT Press), 13–36. doi: 10.7551/mitpress/9780262019279.003.0002
- Unsworth, N., Fukuda, K., Awh, E., and Vogel, E. K. (2014). Working memory and fluid intelligence: capacity, attention control, and secondary memory retrieval. *Cogn. Psychol.* 71, 1–26. doi: 10.1016/j.cogpsych.2014.01.003
- Urgolites, Z. J., and Wood, J. N. (2013). Visual long-term memory stores high-fidelity representations of observed actions. *Psychol. Sci.* 24, 403–411. doi: 10.1177/0956797612457375
- van Boxtel, J. J., and Lu, H. (2015). “Understanding Biological Motion,” in *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, eds N. Pinkerton, R. A. Scott, and S. M. Kosslyn (Hoboken NJ: John Wiley & Sons), 1–14.
- Vanrie, J., and Verfaillie, K. (2004). Perception of biological motion: a stimulus set of human point-light actions. *Behav. Res. Methods Instrum. Comput.* 36, 625–629. doi: 10.3758/bf03206542
- Vicary, S. A., Robbins, R. A., Calvo-Merino, B., and Stevens, C. J. (2014). Recognition of dance-like actions: memory for static posture or dynamic movement? *Mem. Cogn.* 42, 755–767. doi: 10.3758/s13421-014-0395-0
- Vicary, S. A., and Stevens, C. J. (2014). Posture-based processing in visual short-term memory for actions. *Q. J. Exp. Psychol.* 67, 2409–2424. doi: 10.1080/17470218.2014.931445
- Wood, J. N. (2007). Visual working memory for observed actions. *J. Exp. Psychol.* 136, 639–652. doi: 10.1037/0096-3445.136.4.639
- Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition* 108, 522–532. doi: 10.1016/j.cognition.2008.02.012
- Wood, J. N. (2010). Visual working memory retains movement information within an allocentric reference frame. *Vis. Cogn.* 18, 1464–1485. doi: 10.1080/13506285.2010.502430
- Wood, J. N. (2011). A core knowledge architecture of visual working memory. *J. Exp. Psychol.* 37, 357–381. doi: 10.1037/a0021935
- Woodman, G. F., Luck, S. J., and Schall, J. D. (2007). The role of working memory representations in the control of attention. *Cereb. Cortex* 17, 118–124.
- Wu, Y. C., and Coulson, S. (2014). A psychometric measure of working memory capacity for configured body movement. *PLoS One* 9:e84834. doi: 10.1371/journal.pone.0084834
- Xin, F., and Lei, X. (2015). Competition between frontoparietal control and default networks supports social working memory and empathy. *Soc. Cogn. Affect. Neurosci.* 10, 1144–1152. doi: 10.1093/scan/nsu160
- Xu, Y., and Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95. doi: 10.1038/nature04262
- Zhang, W., and Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature* 453, 233–235. doi: 10.1038/nature06860

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ye, Li, Zhang, Gu, Lu, Gao and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Visual Working Memory for Faces and Facial Expressions as a Useful “Tool” for Understanding Social and Affective Cognition

Filippo Gamarota<sup>1</sup> and Paola Sessa<sup>1,2\*</sup>

<sup>1</sup>Department of Developmental Psychology and Socialization, University of Padua, Padua, Italy, <sup>2</sup>Padova Neuroscience Center, University of Padua, Padua, Italy

## OPEN ACCESS

### Edited by:

Zaifeng Gao,  
Zhejiang University, China

### Reviewed by:

Margaret Cecilia Jackson,  
University of Aberdeen,  
United Kingdom  
Xiaomei Zhou,  
Ryerson University, Canada

### \*Correspondence:

Paola Sessa  
paola.sessa@unipd.it

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 19 July 2019

**Accepted:** 07 October 2019

**Published:** 22 October 2019

### Citation:

Gamarota F and Sessa P (2019)  
Visual Working Memory for Faces  
and Facial Expressions as a Useful  
“Tool” for Understanding Social and  
Affective Cognition.  
Front. Psychol. 10:2392.  
doi: 10.3389/fpsyg.2019.02392

Visual working memory (VWM) is one of the most investigated cognitive systems functioning as a *hub* between low- and high-level processes. Remarkably, its role in human cognitive architecture makes it a stage of crucial importance for the study of socio-affective cognition, also in relation with psychopathology such as anxiety. Among socio-affective stimuli, faces occupy a place of first importance. How faces and facial expressions are encoded and maintained in VWM is the focus of this review. Within the main theoretical VWM models, we will review research comparing VWM representations of faces and of other classes of stimuli. We will further present previous work investigating if and how both static (i.e., ethnicity, trustworthiness and identity) and changeable (i.e., facial expressions) facial features are represented in VWM. Finally, we will examine research showing qualitative differences in VWM for face representations as a function of psychopathology and personality traits. The findings that we will review are not always coherent with each other, and for this reason we will highlight the main methodological differences as the main source of inconsistency. Finally, we will provide some suggestions for future research in this field in order to foster our understanding of representation of faces in VWM and its potential role in supporting socio-affective cognition.

**Keywords:** visual working memory, face, facial expressions, social cognition, affective cognition, psychopathology

## INTRODUCTION

Faces are processed in a unique fashion starting from initial perceptual stages (i.e., encoding). The domain-specific approach sustains that face processing is carried out in specialized modules (Kanwisher and Yovel, 2006). Contrarily, the domain-general approach considers common mechanisms that may operate on face and non-facial stimuli as well. In this perspective, the main factor leading to different processing for faces compared to non-facial stimuli is the substantial visual expertise for the former (Gauthier et al., 2000). This debate aside, faces seem to be characterized by distinctive processing from early stages and supported by specific brain areas (Haxby et al., 2000, 2002) that may, at least in part, explain how faces are represented in visual working memory (VWM), also when compared to other non-facial stimuli.

VWM is a core cognitive system defined by a limited-space in terms of capacity in which visual information is temporarily stored and manipulated for further processing (Luck, 2008; Liesefeld and Müller, 2019) and in this vein it can be considered as a “form of mental workspace” (Fukuda et al., 2010).

One important dispute regards VWM storage organization in relation to memory item feature (e.g., semantic category, visual complexity, and expertise). When dealing with visually complex items (like Chinese characters, polygons, and faces) a particular class of models seems relevant. *Flexible resource models* (as opposed to *discrete resolution models*; see Luck et al., 1997; Vogel et al., 2001) propose that a limited pool of memory resources can be allocated in a continuous fashion. Each memory representation has a part of noise and the allocation of a larger amount of memory resources leads to less noise and increases item *resolution*. Memory capacity limit occurs because more complex items require a larger amount of resources compared to simpler items (Alvarez and Cavanagh, 2004; Ma et al., 2014; see also Pratte et al., 2017, for a variant of discrete resolution models that consider systematic variation in precision across the stimuli; see also Swan and Wyble, 2014, for an *hybrid model*; see also van den Berg et al., 2012). Differently, *discrete resolution models* (Luck et al., 1997; Vogel et al., 2001) suggest a fixed slot organization of VWM where each memory item is represented within a slot regardless of the feature complexity. Both approaches consider VWM as characterized by limited capacity (3–4 elements on average); however, the concept of complexity is differently treated. Within flexible resource models, the slope in a visual search rate task (i.e., informational load; Alvarez and Cavanagh, 2004) has been proposed as a quantification of visual complexity. In fact, faces are associated with the slowest search rate (i.e., highest *informational load*) and lowest VWM capacity compared to other stimuli (Eng et al., 2005; Jackson and Raymond, 2008; but see Scolari et al., 2008).

Traditionally, VWM has been studied for simple and abstract stimuli (i.e., colored squares, tilted lines) (Luck et al., 1997; Vogel et al., 2001). Nevertheless, a central aspect of human cognition is the processing of stimuli with social and affective content. To note, according to the importance that VWM may have in social and affective cognition, an updated version of Baddeley's model of working memory (Baddeley and Hitch, 1974) has been more recently proposed considering a specific component devoted to stimuli with emotional content (Baddeley et al., 2012; Xie et al., 2016). Given the importance of VWM in the human cognitive architecture, it is crucial to understand how these emotional stimuli are represented. Among them, faces certainly occupy a place of the highest order. They convey social and affective relevant information such as identity, ethnicity, and emotions.

## METHODOLOGICAL ASPECTS

For a better comprehension of the studies reviewed in the subsequent sections, this section provides a brief overview on methodological aspects related to VWM research.

One of the traditional paradigms to investigate VWM is the *change detection task* (CDT) (Luck et al., 1997; Vogel et al., 2001; Rensink, 2002). Basically, a *memory array* containing to-be-memorized items is presented, and after a blank *retention interval*, a *test display* is displayed. A behavioral response is needed. Participants are required to compare the to-be-memorized items in the memory array with the item/items presented in the test display. These CDT components roughly correspond to the main VWM operations of encoding, maintenance, and retrieval (Luck, 2008; Liesefeld and Müller, 2019). Although other VWM-related paradigms have also been more or less successfully employed, (e.g., the *n-back* task; Jaeggi et al., 2010), the CDT is the most widely used and is considered the most versatile paradigm for the study of VWM (Luck and Vogel, 2013).

Given the extensive use of this paradigm, this has led to a great proliferation of CDT variants, sometimes at the expense of the interpretation of the results. The most common CDT manipulations regard the amount and/or type of the memory array and test display items, the duration of both the memory array (with a significant impact on the amount of available encoding time for each displayed item) and the retention interval, and the type of test display presented after the retention interval (e.g., single probe vs. whole display; see, e.g., Vogel and Machizawa, 2004; Zhang and Luck, 2008; Brigadoi et al., 2017). One important variant regards the use of a continuous probe display (e.g., choice of a to-be-remembered color from a colors wheel) allowing an estimation of memory precision (Zhang and Luck, 2008; see also Lorenc et al., 2014; Krill et al., 2018 for examples with faces). Other possible variants concern the use of distractors or masks during the retention interval (Vogel et al., 2006).

Within the context of studies that used the CDT, several VWM-dependent measures have been used, including measures of storage capacity (e.g., Cowan's *K*; Cowan, 2001) – an index of the amount of items effectively retained (for a review on capacity measures, see Rouder et al., 2011) – measures of accuracy – the percentage of correct responses – and measures of sensitivity in the comparison task between the to-be-memorized items and that/those presented in the test display (e.g., *d'* from signal detection theory; Green and Swets, 1974; Wilken and Ma, 2004). As mentioned before, a continuous probe display allows the memory precision estimation through an error distribution around the right value. Finally, the concept of *informational load* (Alvarez and Cavanagh, 2004; Eng et al., 2005) mentioned above is frequently used to compare different stimuli with regard to their visual complexity (but see Jiang et al., 2008).

One of the most studied neural correlate of VWM is an event-related potential (ERP) called *contralateral delay activity* (CDA) or also *sustained posterior contralateral negativity* (SPCN) (for a review, see Luria et al., 2016). This ERP is recorded at occipito-parietal electrodes (*ibidem*) and it has been suggested that the intraparietal sulcus (IPS) is the main neural generator (Xu and Chun, 2006; Robitaille et al., 2009). It is computed as a difference wave (Gratton, 1998) between contralateral and ipsilateral activity related to the hemifield location of to-be-memorized items. CDA amplitude tends to correlate

with the amount (Vogel and Machizawa, 2004) and resolution (Luria et al., 2016) of stored visual information and it is also sensitive to visual complexity (colors vs. random polygons; Luria et al., 2010).

Given the great variability in the methods employed and results obtained in the context of VWM studies, we selected those investigations that used a comparable methodology in order to facilitate comparison between results. In some cases, the results of the different studies discussed here are not directly comparable because of differences in the stimuli used (e.g., schematic faces vs. real faces, different facial expressions, etc.) and/or participants' task (detection of a change in faces identity vs. facial expressions). For this reason, we have tried to indicate details useful to the readers for a critical analysis of the results. In the following sections, we focus on studies using CDT with faces with particular attention to studies that measured the CDA. In the last section of this review, we also discuss studies that considered the relationship between face representations and individual differences (e.g., psychopathology). This review does not aim to be exhaustive but rather to identify and present selected examples of evidence that may help clarify the critical link between VWM functioning and the complexity of social cognition focusing on the main source of social information, that is others' faces.

## FACES AND VISUAL WORKING MEMORY

Curby and Gauthier (2007) demonstrated that a greater number of upright stimuli can be retained in VWM (measured with Cowan's  $K$ ) compared to inverted ones, and, according to the *face inversion* effect (Yin, 1969; Tanaka and Gordon, 2011), this effect is larger for faces compared to non-facial stimuli (for a review see McKone and Robbins, 2011). Also, the precision is higher for upright faces when compared to inverted faces (Lorenc et al., 2014; Krill et al., 2018). Furthermore, coherent with face visual complexity (Eng et al., 2005; Jiang et al., 2008), this effect is present only if sufficient encoding time (i.e., memory array duration) is provided. One possible explanation for this pattern of results takes into account holistic/configural processing that characterizes faces. In support of this, similar VWM advantage has been reported in expert individuals with other class of objects (Curby et al., 2009; but see Wong et al., 2008; Jiang et al., 2016) or famous faces (Jackson and Raymond, 2008). Within the theoretical framework considering the dissociation between capacity, in terms of slots, and resolution of VWM representations (Scolari et al., 2008; Zhang and Luck, 2008), it has been suggested that perceptual expertise may enhance the resolution of VWM representations (Scolari et al., 2008; Curby and Gauthier, 2010; Lorenc et al., 2014). These results are noteworthy as they strongly suggest that resolution may be a particularly flexible aspect of VWM and potentially modulated on the basis of factors such as, in this case, perceptual expertise, but possibly also social and emotional salience. Therefore, VWM resolution could be a key element for

understanding VWM representations of faces and facial expressions of emotions.

## Static and Changeable Facial Features

Faces are characterized by both static and changeable features that convey social and affective information, such as race, identity, trustworthiness (Oosterhof and Todorov, 2009), facial expressions, and gaze direction (Adolphs and Birmingham, 2011).

Recognizing people's identity is a fundamental social ability (Bruce and Young, 1986; Haxby et al., 2000) and it has been suggested familiarity with specific individual faces might affect their storage in VWM. For this reason, face familiarity could influence VWM in real-time identity processing. Jackson and Raymond (2008) using an identity CDT demonstrated a VWM improvement (capacity and sensitivity) for familiar actors' faces compared to unfamiliar ones, leading to the conclusion of an involvement of long-term memory in VWM representations of familiar faces. The effect disappeared for inverted faces. Testing pictorial details' representations of different pictures of the same individual – either familiar or unfamiliar – Dunn et al. (2019, see exp. 2–3) found no difference in performance (in terms of sensitivity) as a function of familiarity.

Race seems to influence the quality of face processing (Young et al., 2012) possibly influencing VWM representations. Zhou et al. (2018) demonstrated that with short encoding time, other-race faces are retained with reduced precision (i.e., standard deviation of errors distributions) compared to own-race faces. Stelter and Degner (2018) demonstrated both lower accuracy ( $d'$ ) and capacity (Cowan's  $K$ ) for other-race faces. These findings suggest that, similar to inverted faces, other-race faces are processed, at both configural and featural levels of processing, less efficiently (Hayward et al., 2013; Stelter and Degner, 2018). Holistic/configural processing seems a critical aspect in race processing (Tanaka et al., 2004), that may also depend on other social-cognitive factors linked to intergroup processing (for a review, see Young et al., 2012). Interestingly, a previous study has also provided evidence of a reduced CDA amplitude for other-race faces, especially with direct gaze (Sessa and Dalmaso, 2016) and another study reported a correlation between CDA amplitude and implicit racial prejudice scores (Sessa et al., 2012), such that the most prejudiced participants memorized other-race faces with the lowest resolution.

Facial expressions are extremely relevant to social cognition. Information on the others' affective states (e.g., others' emotions) and on the environment (e.g., dangers from fearful reactions) could be extracted from facial expressions (Adolphs, 2002). Using similar methodology (i.e., a single-probe identity CDT with real faces; facial expression was task-irrelevant), one recurring finding in VWM literature is that of an advantage in terms of capacity (Cowan's  $K$ ) and sensitivity ( $d'$ ) for negative facial expressions, especially angry, compared to happy and neutral expressions (Jackson et al., 2008, 2009, 2014; Thomas et al., 2014).

Furthermore, this benefit is observed only when angry faces are presented in the memory array but not in the test



display (Jackson et al., 2014). In addition, it declines during the retention interval. Using a longer retention interval (i.e., 9,000 vs. 1,000 ms in the study by Jackson et al., 2009) this benefit disappears (Jackson et al., 2012). Notably, this angry benefit occurs without reducing performance for concurrently presented neutral faces. All stimuli are retained, with an increased resolution for salient stimuli (Thomas et al., 2014). However, slightly different results (i.e., the absence of an angry benefit and/or the presence of a happy benefit) have been reported using schematic facial expressions (i.e., no information on identity), shorter encoding times, or other different methodological details (Langeslag et al., 2009; Simione et al., 2014; Xie et al., 2016; Spotorno et al., 2018; Curby et al., 2019). In particular, the angry face advantage has not always been observed (see also Curby et al., 2019 using a change localization task; Xie et al., 2016 using schematic faces) or has been reported only for short encoding times (150 vs. 1,000/2,000 ms of the previously cited studies) (Simione et al., 2014 using schematic faces).

Varying memory array size, encoding time, and expression (fearful, happy, angry, and neutral), Curby et al. (2019) demonstrated a VWM “cost” for fearful, compared to neutral and happy real faces in terms of lower capacity (Cowan’s  $K$ ). Opposite to the angry benefit, a cost for angry faces has been also observed (Curby et al., 2019) when compared to happy faces (indeed a happy benefit emerged). To note, other studies have instead demonstrated a fearful advantage in terms of capacity, accuracy, and CDA amplitude (Sessa et al., 2011; Stout et al., 2013; Lee and Cho, 2019; all studies used real faces and facial expression was task-irrelevant). Methodological differences could at least in part explain these inconsistent findings. Sessa et al. (2011) and Stout et al. (2013) used a shorter encoding time (200–500) and a smaller set size (1–2) when compared to the study by (Curby et al., 2019; 1,000/4,000 ms and five items, respectively) and the spatial information was less relevant (i.e., the location was probed in Curby et al., 2019). Interestingly, in Curby et al.’s (2019) study, the fear cost emerged only at the longest encoding time and, as argued by authors, a difficulty in disengaging from fearful faces could explain the lower estimated capacity. When controlling for spatial and temporal attention, a fearful advantage in terms of sensitivity ( $d'$ ) emerges (Lee and Cho, 2019).

Overall, the angry face benefit seems consistent across studies. However, changing some CDT parameters like probing method (i.e., probed location), using real vs. schematic faces, different encoding times and/or dependent variables (Cowan  $K$  vs.  $d'$ ) seems to influence this effect (Langeslag et al., 2009; Simione et al., 2014; Xie et al., 2016; Spotorno et al., 2018; Curby et al., 2019). Similarly, a fearful advantage, relative to neutral faces, is observed for studies using similar parameters (Sessa et al., 2011; Stout et al., 2013; Lee and Cho, 2019; but see Curby et al., 2019). Importantly, CDA seems to differentiate fearful and neutral faces regardless of set size and spatial or temporal attentional biases (Sessa et al., 2011) and this may indicate that, compared to VWM

behavioral estimates, CDA is more sensitive to resolution variations according to saliency.

## Other Socially Relevant Factors and Interindividual Differences

Other investigations combined different emotional stimuli for understanding how social information is integrated into VWM. Negative emotional words presented during the retention interval (2,000 ms) seem to enhance performance ( $d'$ ) for angry faces (compared to happy) (Jackson et al., 2014). An angry benefit emerged with both positive and negative words when using a longer retention interval (9,000 ms; Jackson et al., 2012). Authors suggested that encoding negative faces creates a condition (*threat tagging*) in which identity is coupled with valence and congruent stimuli (i.e., negative words) can interact with this representation (Jackson et al., 2014). Maran et al. (2015) induced positive or negative mood using high-impact pictures (e.g., erotic, mutilations, etc.) and observed improved performance ( $d'$ ) for all emotional faces. Similarly, inducing a feeling of social exclusion (Du et al., 2019) or including a monetary reward (instead of penalty; Thomas et al., 2016) improved VWM capacity for faces. On the contrary, a facial task during the retention interval while maintaining a face in VWM seems to decrease accuracy (Robinson et al., 2008). Overall, VWM for faces seems to benefit from non-facial emotional stimuli such as words or other non-visual factors (i.e., mood).

Dealing with task-relevant and irrelevant (distractors) information is another important VWM facet. Filtering efficiency interacts with individual VWM capacity (Vogel et al., 2005) and with psychopathology (Stout et al., 2013, 2015). CDA seems to be an optimal measure for this purpose. Given the correlation with the number of to-be-memorized items until capacity limit (Vogel and Machizawa, 2004), CDA amplitude for  $n$  task-relevant stimuli should be greater than amplitude for  $n$  stimuli, some of which are task-irrelevant. Including emotional face distractors in the memory array (happy, angry, and neutral) and using an identity CDT (1 or 2 to-be-remembered faces), Ye et al. (2018) found that high-capacity subjects filtered out all distractors compared to low-capacity subjects in whom filtering activity was effective only for happy faces.

Psychopathology is another critical factor in social cognition. Anxiety, in particular, has been widely studied in relation to WM and generally correlates with lower WM capacity (for a review, see Moran, 2016). In two different experiments using a location probe task with real emotional faces (angry, neutral, and happy), Yao et al. (2018) demonstrated lower VWM capacity (Cowan’s  $K$ ) for all facial expressions in individuals with higher self-reported anxiety, without affecting precision.

Filtering irrelevant information is an important WM function and could be relevant in anxiety (Qi et al., 2014). Using an identity CDT and monitoring the CDA, Stout et al. (2013) measured the filtering efficiency for task-irrelevant faces (with fearful or neutral expressions). They found that task-irrelevant fearful faces were less efficiently filtered out compared to neutral faces. In addition, filtering efficiency negatively correlated with self-reported anxiety. More specifically, Stout et al. (2015)



demonstrated that filtering efficiency is specifically inversely related to the worry component of anxiety. Moreover, Meconi et al. (2013) using an identity CDT reported greater CDA amplitude for trustworthy faces. Interestingly, when self-reported anxiety was considered, untrustworthy faces (vs. trustworthy) were associated with larger CDA amplitude in the most anxious participants.

Other clinical conditions have been studied in relation to facial expression VWM representations. Patients with schizophrenia seem to have an overall WM deficit (Forbes et al., 2009) and lower VWM capacity for neutral faces (She et al., 2019). Interestingly, using emotional faces, the angry benefit is still present although an emotion classification deficit is observed (Linden et al., 2010). Individuals with melancholic depression have a VWM bias (i.e., higher  $d'$ ) toward sad faces compared to individuals with non-melancholic depression (Linden et al., 2011). In an expression change localization task, individuals with high suicidal intentions seem to have worse VWM capacity for negative schematic faces compared to controls (Xie et al., 2018). Furthermore, Takahashi et al. (2015) using a CDT with schematic faces (angry, happy, and neutral) demonstrated that high alexithymic individuals have worse VWM capacity for happy faces compared to low alexithymic individuals.

## DISCUSSION AND CONCLUSION

Faces are complex stimuli that convey multiple information and that seem to be subject to a special type of holistic processing during early stages of processing. For this reason, it is plausible to hypothesize that faces are also represented in VWM in a “special” way when compared to non-facial stimuli or inverted faces. Many of the studies in the literature have focused on the effects of facial expressions of emotions (both task-relevant with schematic faces, and task-irrelevant with real faces of different identities) on the representation of faces in VWM. Negative faces, in particular angry, are associated with better VWM performances. However, a great methodological variability in stimuli choice and CDT parameters makes it difficult to compare findings. As previously shown, results could drastically change using schematic vs. real faces or different probing methods. Future research in this field, if not of interest, should keep paradigms’ parameters fixed,

only varying socially relevant information. Otherwise, an orthogonal variation of CDT parameters within the same study could be useful (e.g., using several encoding times, schematic vs. real faces).

VWM is defined a hub of cognition (Luck, 2008) where information is retained and manipulated. Interestingly, different socially relevant information (e.g., emotional words or mood) seems to interact with facial memory representations. Ecologically, integrating different sources of social information could be an adaptive mechanism.

Psychopathology is another important aspect in social environment and often related to changes in basic cognitive functions. Again, different methods and different psychopathological conditions are difficult to integrate. However, it is interesting noting that psychopathology and VWM functioning are related. Alexithymic individuals have the worst VWM performance for happy faces (Takahashi et al., 2015) and individuals with suicidal intentions show the worst VWM performance for negative stimuli, probably originating from an adaptive avoidance behavior (Xie et al., 2018).

At the neural level, the CDA seems to be influenced by facial information. It has been demonstrated that the CDA is modulated according to the amount (Vogel and Machizawa, 2004) and also the quality (i.e., resolution) of visual information (Luria et al., 2016). Interestingly, even with a single to-be-remembered face (i.e., capacity estimation is not relevant), the CDA is modulated by facial information (Sessa et al., 2011, 2018; Meconi et al., 2013). According to *flexible resource models* and the *neural object-file theory* (Xu and Chun, 2006, 2009), one important and ecologically relevant aspect to be considered could be the resolution variation according to saliency. The theory proposes two stages of processing (with neural bases on distinct part of IPS that is supposed to be also the CDA generator), where the second stage regards a detailed visual encoding of relevant objects. Integrating this neural measure in standard behavioral studies and focusing on resolution besides capacity could be useful for finely comparing representations of different socially relevant information.

## AUTHOR CONTRIBUTIONS

FG wrote the first draft of the manuscript. PS provided critical revision. Both authors read and approved the submitted version.

## REFERENCES

- Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav. Cogn. Neurosci. Rev.* 1, 21–62. doi: 10.1177/1534582302001001003
- Adolphs, R., and Birmingham, E. (2011). “Neural substrates of social perception” in *Oxford handbook of face perception*. eds. A. J. Calder, G. Rodhe, M. H. Johnson, and J. V. Haxby (Oxford: Oxford University Press), 571–589.
- Alvarez, G. A. A., and Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. Sci.* 15, 106–111. doi: 10.1111/j.0963-7214.2004.01502006.x
- Baddeley, A., Banse, R., Huang, Y. M., and Page, M. (2012). Working memory and emotion: detecting the hedonic detector. *J. Cogn. Psychol.* 24, 6–16. doi: 10.1080/20445911.2011.613820
- Baddeley, A., and Hitch, G. (1974). “Working memory” in *The psychology of learning and motivation*. ed. G. H. Bower (New York: Academic Press), 47–89.
- Brigadoi, S., Cutini, S., Meconi, F., Castellaro, M., Sessa, P., Marangon, M., et al. (2017). On the role of the inferior intraparietal sulcus in visual working memory for lateralized single-feature objects. *J. Cogn. Neurosci.* 29, 337–351. doi: 10.1162/jocn\_a\_01042
- Bruce, V., and Young, A. (1986). Understanding face recognition. *Br. J. Psychol.* 77, 305–327. doi: 10.1111/j.2044-8295.1986.tb02199.x

- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. doi: 10.1017/S0140525X01003922
- Curby, K. M., and Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychon. Bull. Rev.* 14, 620–628. doi: 10.3758/BF03196811
- Curby, K. M., and Gauthier, I. (2010). To the trained eye: perceptual expertise alters visual processing. *Top. Cogn. Sci.* 2, 189–201. doi: 10.1111/j.1756-8765.2009.01058.x
- Curby, K. M., Glazek, K., and Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 94–107. doi: 10.1037/0096-1523.35.1.94
- Curby, K. M., Smith, S. D., Moerel, D., and Dyson, A. (2019). The cost of facing fear: visual working memory is impaired for faces expressing fear. *Br. J. Psychol.* 110, 428–448. doi: 10.1111/bjop.12324
- Du, X., Xu, M., Ding, C., Yuan, S., Zhang, L., and Yang, D. (2019). Social exclusion increases the visual working memory capacity of social stimuli. *Curr. Psychol.* 1–12. doi: 10.1007/s12144-019-00274-1
- Dunn, J. D., Ritchie, K. L., Kemp, R. I., and White, D. (2019). Familiarity does not inhibit image-specific encoding of faces. *J. Exp. Psychol. Hum. Percept. Perform.* 45, 841–854. doi: 10.1037/xhp0000625
- Eng, H. Y., Chen, D., and Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychon. Bull. Rev.* 12, 1127–1133. doi: 10.3758/BF03206454
- Forbes, N. F., Carrick, L. A., McIntosh, A. M., and Lawrie, S. M. (2009). Working memory in schizophrenia: a meta-analysis. *Psychol. Med.* 39, 889–905. doi: 10.1017/S0033291708004558
- Fukuda, K., Awh, E., and Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Curr. Opin. Neurobiol.* 20, 177–182. doi: 10.1016/j.conb.2010.03.005
- Gauthier, I., Skudlarski, P., Gore, J. C., Anderson, A. W., Skudlarsky, P., Gore, J. C., et al. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci.* 3, 191–197. doi: 10.1038/72140
- Gratton, G. (1998). The contralateral organization of visual memory: a theoretical concept and a research tool. *Psychophysiology* 35, 638–647. doi: 10.1111/1469-8986.3560638
- Green, D. M., and Swets, J. A. (1974). *Signal detection theory and psychophysics*. Oxford, England: Robert E. Krieger.
- Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn. Sci.* 4, 223–233. doi: 10.1016/S1364-6613(00)01482-0
- Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol. Psychiatry* 51, 59–67. doi: 10.1016/S0006-3223(01)01330-0
- Hayward, W. G., Crookes, K., and Rhodes, G. (2013). The other-race effect: holistic coding differences and beyond. *Vis. Cogn.* 21, 1224–1247. doi: 10.1080/13506285.2013.824530
- Jackson, M. C., Linden, D. E. J., and Raymond, J. E. (2012). “Distracters” do not always distract: visual working memory for angry faces is enhanced by incidental emotional words. *Front. Psychol.* 3:437. doi: 10.3389/fpsyg.2012.00437
- Jackson, M. C., Linden, D. E. J., and Raymond, J. E. (2014). Angry expressions strengthen the encoding and maintenance of face identity representations in visual working memory. *Cogn. Emot.* 28, 278–297. doi: 10.1080/02699931.2013.816655
- Jackson, M. C., and Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 556–568. doi: 10.1037/0096-1523.34.3.556
- Jackson, M. C., Wolf, C., Johnston, S. J., Raymond, J. E., and Linden, D. E. J. (2008). Neural correlates of enhanced visual short-term memory for angry faces: an fMRI study. *PLoS One* 3:e3536. doi: 10.1371/journal.pone.0003536
- Jackson, M. C., Wu, C.-Y., Linden, D. E. J., and Raymond, J. E. (2009). Enhanced visual short-term memory for angry faces. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 363–374. doi: 10.1037/a0013895
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., and Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory* 18, 394–412. doi: 10.1080/09658211003702171
- Jiang, Y. V., Remington, R. W., Asaad, A., Lee, H. J., and Mikkalson, T. C. (2016). Remembering faces and scenes: the mixed-category advantage in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1399–1411. doi: 10.1037/xhp0000228
- Jiang, Y. V., Shim, W. M., and Makovski, T. (2008). Visual working memory for line orientations and face identities. *Atten. Percept. Psychophys.* 70, 1581–1591. doi: 10.3758/PP.70.8.1581
- Kanwisher, N., and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philos. Trans. R. Soc. B Biol. Sci.* 361, 2109–2128. doi: 10.1098/rstb.2006.1934
- Krill, D., Avidan, G., and Pertzov, Y. (2018). The rapid forgetting of faces. *Front. Psychol.* 9, 1–13. doi: 10.3389/fpsyg.2018.01319
- Langeslag, S. J. E., Morgan, H. M., Jackson, M. C., Linden, D. E. J., and Van Strien, J. W. (2009). Electrophysiological correlates of improved short-term memory for emotional faces. *Neuropsychologia* 47, 887–896. doi: 10.1016/j.neuropsychologia.2008.12.024
- Lee, H. J., and Cho, Y. S. (2019). Memory facilitation for emotional faces: visual working memory trade-offs resulting from attentional preference for emotional facial expressions. *Mem. Cogn.* 1231–1243. doi: 10.3758/s13421-019-00930-8
- Liesefeld, H. R., and Müller, H. J. (2019). Current directions in visual working memory research: an introduction and emerging insights. *Br. J. Psychol.* 110, 193–206. doi: 10.1111/bjop.12377
- Linden, S. C., Jackson, M. C., Subramanian, L., Healy, D., and Linden, D. E. J. (2011). Sad benefit in face working memory: an emotional bias of melancholic depression. *J. Affect. Disord.* 135, 251–257. doi: 10.1016/j.jad.2011.08.002
- Linden, S. C., Jackson, M. C., Subramanian, L., Wolf, C., Green, P., Healy, D., et al. (2010). Emotion-cognition interactions in schizophrenia: implicit and explicit effects of facial expression. *Neuropsychologia* 48, 997–1002. doi: 10.1016/j.neuropsychologia.2009.11.023
- Lorenc, E. S., Pratte, M. S., Angeloni, C. F., and Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Atten. Percept. Psychophys.* 76, 1975–1984. doi: 10.3758/s13414-014-0653-z
- Luck, S. J. (2008). “Visual short-term memory” in *Visual memory*. eds. S. J. Luck, and A. Hollingworth (New York, NY: Oxford University Press), 43–85.
- Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006
- Luck, S. J., Vogel, E. K., Vogel, J., and Edward, K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–284. doi: 10.1038/36846
- Luria, R., Balaban, H., Awh, E., and Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neurosci. Biobehav. Rev.* 62, 100–108. doi: 10.1016/j.neubiorev.2016.01.003
- Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., Dell’Acqua, R., Jolicoeur, P., et al. (2010). Visual short-term memory capacity for simple and complex objects. *J. Cogn. Neurosci.* 22, 496–512. doi: 10.1162/jocn.2009.21214
- Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356. doi: 10.1038/nn.3655
- Maran, T., Sachse, P., and Furtner, M. (2015). From specificity to sensitivity: affective states modulate visual working memory for emotional expressive faces. *Front. Psychol.* 6:1297. doi: 10.3389/fpsyg.2015.01297
- McKone, E., and Robbins, R. (2011). “Are faces special?” in *Oxford handbook of face perception*. eds. A. J. Calder, G. Rhodes, M. H. Johnson, and J. V. Haxby (Oxford, UK: Oxford University Press).
- Meconi, F., Luria, R., and Sessa, P. (2013). Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. *Soc. Cogn. Affect. Neurosci.* 9, 1872–1879. doi: 10.1093/scan/nst189
- Moran, T. P. (2016). Anxiety and working memory capacity: a meta-analysis and narrative review. *Psychol. Bull.* 142, 831–864. doi: 10.1037/bul0000051
- Oosterhof, N. N., and Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion* 9, 128–133. doi: 10.1037/a0014520
- Pratte, M. S., Park, Y. E., Rademaker, R. L., and Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 6–17. doi: 10.1037/xhp0000302
- Qi, S., Ding, C., and Li, H. (2014). Neural correlates of inefficient filtering of emotionally neutral distractors from working memory in trait anxiety. *Cogn. Affect. Behav. Neurosci.* 14, 253–265. doi: 10.3758/s13415-013-0203-5
- Rensink, R. A. (2002). Change detection. *Annu. Rev. Psychol.* 53, 245–277. doi: 10.1146/annurev.psych.53.100901.135125

- Robinson, A., Manzi, A., and Triesch, J. (2008). Object perception is selectively slowed by a visually similar working memory load. *J. Vis.* 8, 1–13. doi: 10.1167/8.16.7
- Robitaille, N., Grimault, S., and Jolicoeur, P. (2009). Bilateral parietal and contralateral responses during maintenance of unilaterally encoded objects in visual short-term memory: evidence from magnetoencephalography. *Psychophysiology* 46, 1090–1099. doi: 10.1111/j.1469-8986.2009.00837.x
- Rouder, J. N., Morey, R. D., Morey, C. C., and Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychon. Bull. Rev.* 18, 324–330. doi: 10.3758/s13423-011-0055-3
- Scolari, M., Vogel, E. K., and Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychon. Bull. Rev.* 15, 215–222. doi: 10.3758/PBR.15.1.215
- Sessa, P., and Dalmasso, M. (2016). Race perception and gaze direction differently impair visual working memory for faces: an event-related potential study. *Soc. Neurosci.* 11, 97–107. doi: 10.1080/17470919.2015.1040556
- Sessa, P., Jolicoeur, P., Dell'Acqua, R., Gotler, A., Luria, R., Gotler, A., et al. (2011). Interhemispheric ERP asymmetries over inferior parietal cortex reveal differential visual working memory maintenance for fearful versus neutral facial identities. *Psychophysiology* 48, 187–197. doi: 10.1111/j.1469-8986.2010.01046.x
- Sessa, P., Schiano Lomoriello, A., and Luria, R. (2018). Neural measures of the causal role of observers' facial mimicry on visual working memory for facial expressions. *Soc. Cogn. Affect. Neurosci.* 13, 1281–1291. doi: 10.1093/scan/nsy095
- Sessa, P., Tomelleri, S., Luria, R., Castelli, L., Reynolds, M., and Dell'Acqua, R. (2012). Look out for strangers! Sustained neural activity during visual working memory maintenance of other-race faces is modulated by implicit racial prejudice. *Soc. Cogn. Affect. Neurosci.* 7, 314–321. doi: 10.1093/scan/nsr011
- She, S., Zhang, B., Mi, L., Li, H., Kuang, Q., Bi, T., et al. (2019). Stimuli may have little impact on the deficit of visual working memory accuracy in first-episode schizophrenia. *Neuropsychiatr. Dis. Treat.* 15, 481–489. doi: 10.2147/NDT.S188645
- Simione, L., Calabrese, L., Marucci, F. S., Belardinelli, M. O., Raffone, A., and Maratos, F. A. (2014). Emotion based attentional priority for storage in visual short-term memory. *PLoS One* 9:e95261. doi: 10.1371/journal.pone.0095261
- Spotorno, S., Evans, M., and Jackson, M. C. (2018). Remembering who was where: a happy expression advantage for face identity-location binding in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 44, 1365–1383. doi: 10.1037/xlm0000522
- Stelter, M., and Degner, J. (2018). Investigating the other-race effect in working memory. *Br. J. Psychol.* 109, 777–798. doi: 10.1111/bjop.12304
- Stout, D. M., Shackman, A. J., Johnson, J. S., and Larson, C. L. (2015). Worry is associated with impaired gating of threat from working memory. *Emotion* 15, 6–11. doi: 10.1037/emo0000015
- Stout, D. M., Shackman, A. J., and Larson, C. L. (2013). Failure to filter: anxious individuals show inefficient gating of threat from working memory. *Front. Hum. Neurosci.* 7:58. doi: 10.3389/fnhum.2013.00058
- Swan, G., and Wyble, B. (2014). The binding pool: a model of shared neural resources for distinct items in visual working memory. *Atten. Percept. Psychophys.* 76, 2136–2157. doi: 10.3758/s13414-014-0633-3
- Takahashi, J., Hirano, T., and Gyoba, J. (2015). Effects of facial expressions on visual short-term memory in relation to alexithymia traits. *Pers. Individ. Dif.* 83, 128–135. doi: 10.1016/j.paid.2015.04.010
- Tanaka, J. W., and Gordon, I. (2011). "Features, configuration, and holistic face processing" in *The Oxford handbook of face perception*. eds. A. J. Calder, G. Rodhe, M. H. Johnson, and J. V. Haxby (New York, NY: Oxford University Press), 177–194.
- Tanaka, J. W., Kiefer, M., and Bukach, C. M. (2004). A holistic account of the own-race effect in face recognition: evidence from a cross-cultural study. *Cognition* 93, B1–B9. doi: 10.1016/j.cognition.2003.09.011
- Thomas, P. M. J., FitzGibbon, L., and Raymond, J. E. (2016). Value conditioning modulates visual working memory processes. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 6–10. doi: 10.1037/xhp0000144
- Thomas, P. M. J., Jackson, M. C., and Raymond, J. E. (2014). A threatening face in the crowd: effects of emotional singletons on visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 253–263. doi: 10.1037/a0033970
- van den Berg, R., Shin, H., Chou, W. C., George, R., and Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proc. Natl. Acad. Sci. USA* 109, 8780–8785. doi: 10.1073/pnas.1117465109
- Vogel, E. K., and Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748–751. doi: 10.1038/nature02447
- Vogel, E. K., McCollough, A. W., and Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature* 438, 500–503. doi: 10.1038/nature04171
- Vogel, E. K., Woodman, G. F., and Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 92–114. doi: 10.1037/0096-1523.27.1.92
- Vogel, E. K., Woodman, G. F., and Luck, S. J. (2006). The time course of consolidation in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1436–1451. doi: 10.1037/0096-1523.32.6.1436
- Wilken, P., and Ma, W. J. (2004). A detection theory account of change detection. *J. Vis.* 4, 1120–1135. doi: 10.1167/4.12.11
- Wong, J. H., Peterson, M. S., and Thompson, J. C. (2008). Visual working memory capacity for objects from different categories: a face-specific maintenance effect. *Cognition* 108, 719–731. doi: 10.1016/j.cognition.2008.06.006
- Xie, W., Li, H., Ying, X., Zhu, S., Fu, R., Zou, Y., et al. (2016). Affective bias in visual working memory is associated with capacity. *Cogn. Emot.* 31, 1–16. doi: 10.1080/02699931.2016.1223020
- Xie, W., Li, H., Zou, Y., Sun, X., and Shi, C. (2018). A suicidal mind tends to maintain less negative information in visual working memory. *Psychiatry Res.* 262, 549–557. doi: 10.1016/j.psychres.2017.09.043
- Xu, Y., and Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95. doi: 10.1038/nature04262
- Xu, Y., and Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends Cogn. Sci.* 13, 167–174. doi: 10.1016/j.tics.2009.01.008
- Yao, N., Chen, S., and Qian, M. (2018). Trait anxiety is associated with a decreased visual working memory capacity for faces. *Psychiatry Res.* 270, 474–482. doi: 10.1016/j.psychres.2018.10.018
- Ye, C., Xu, Q., Liu, Q., Cong, F., Saariluoma, P., Ristaniemi, T., et al. (2018). The impact of visual working memory capacity on the filtering efficiency of emotional face distractors. *Biol. Psychol.* 138, 63–72. doi: 10.1016/j.biopsycho.2018.08.009
- Yin, R. K. (1969). Looking at upside-down faces. *J. Exp. Psychol.* 81, 141–145. doi: 10.1037/h0027474
- Young, S. G., Hugenberg, K., Bernstein, M. J., and Sacco, D. F. (2012). Perception and motivation in face recognition: a critical review of theories of the cross-race effect. *Personal. Soc. Psychol. Rev.* 16, 116–142. doi: 10.1177/1088868311418987
- Zhang, W., and Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature* 453, 233–235. doi: 10.1038/nature06860
- Zhou, X., Mondloch, C. J., and Emrich, S. M. (2018). Encoding differences affect the number and precision of own-race versus other-race faces stored in visual working memory. *Atten. Percept. Psychophys.* 80, 702–712. doi: 10.3758/s13414-017-1467-6

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Gambarota and Sessa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Persistence and Accumulation of Visual Memories for Objects in Scenes in 12-Month-Old Infants

Sylvia B. Guillory<sup>1\*</sup> and Zsuzsa Kaldy<sup>2</sup>

<sup>1</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States, <sup>2</sup>Psychology Department, University of Massachusetts Boston, Boston, MA, United States

Visual memory for objects has been studied extensively in infants over the past 20 years, however, little is known about how they are formed when objects are embedded in naturalistic scenes. In adults, memory for objects in a scene show information *accumulation* over time as well as *persistence* despite interruptions (Melcher, 2001, 2006). In the present study, eye-tracking was used to investigate these two processes in 12-month-old infants ( $N = 19$ ) measuring: (1) whether longer encoding time can improve memory performance (accumulation), and (2) whether multiple shorter exposures to a scene are equivalent to a single exposure of the same total duration (persistence). A control group of adults was also tested in a closely matched paradigm ( $N = 23$ ). We found that increasing exposure time led to gains in memory performance in both groups. Infants were found to be successful in remembering objects with continuous exposures to a scene, but unlike adults, were not able to perform better than chance when interrupted. However, infants' scan patterns showed evidence of memory as they continued the exploration of the scene in a strategic way following the interruption. Our findings provide insight into how infants are able to build representations of their visual environment by accumulating information about objects embedded in scenes.

**Keywords:** visual memory, infants, encoding, persistence, accumulation, interruption, objects, scenes

## OPEN ACCESS

### Edited by:

Hagit Magen,  
Hebrew University of Jerusalem, Israel

### Reviewed by:

Andrea Helo,  
University of Chile, Chile  
Koleen McCrink,  
Columbia University, United States

### \*Correspondence:

Sylvia B. Guillory  
sylvia.guillory@mssm.edu

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 12 December 2018

**Accepted:** 16 October 2019

**Published:** 06 November 2019

### Citation:

Guillory SB and Kaldy Z (2019)  
Persistence and Accumulation of  
Visual Memories for Objects in  
Scenes in 12-Month-Old Infants.  
Front. Psychol. 10:2454.  
doi: 10.3389/fpsyg.2019.02454

Natural scenes are semantically coherent images of a real-world environment comprising of background elements (typically larger scale surfaces, such as ground, walls, and floors) and multiple distinct objects (smaller scale entities, such as plants, cars, and chairs) (Henderson and Hollingworth, 1998). Semantic cohesion and regularities aid visual memory performance (Hollingworth and Henderson, 2000; Brady et al., 2009a) and provide contextual cues (Torralba et al., 2006). Visual memory is necessary to accumulate information obtained from the different fixations as the eyes scan the environment (Melcher, 2001; Hollingworth, 2004). This process requires building a complex representation that contains objects that are bound to locations in the scene's spatial layout and stored in memory (Hollingworth, 2007). The current study investigated how such representations of the visual environment are constructed and maintained in infants.

A fundamental characteristic of visual working memory (VWM) is its limited capacity. Luck and Vogel (1997) found that for adults, VWM stores approximately four units of information. Using a change detection task, adult participants were shown a set of simple objects, such as colored squares; then after a brief delay, the test display was presented where one of the squares may have changed in color. Participants were instructed to indicate whether the two displays were the same or different. These results and similar studies suggest an upper limit to the number of items that can be individuated and maintained (Cowan, 2001, 2010; Scholl and Xu, 2001; Vogel et al., 2001; Alvarez and Cavanagh, 2004; Awh et al., 2007). How much information



can be actively stored in VWM has significant consequences on learning and other adaptive functioning (Fukuda et al., 2010).

In development, an emerging picture reveals a gradual increase in VWM capacity over the first year (Rose et al., 2001; Kaldy and Leslie, 2003, 2005; Oakes et al., 2006, 2017; Kibbe and Leslie, 2011; Kwon et al., 2014; Kaldy et al., 2016) that continues to develop into childhood (Simmering, 2012; Guillory et al., 2018). Using a version of the change detection task with three objects, Oakes et al. (2006) found that 8-month-old infants succeeded at binding objects to their locations. Kaldy and Leslie (2003, 2005) reported that 9-month-old, but not 6-month-old, infants looked longer when two objects unexpectedly switched locations. VWM capacity has also been studied in older infants using a manual search paradigm. In these studies, objects are placed into an opaque box and infants are later given the opportunity to search the box and retrieve the hidden objects. Results show that 10- and 12-month-olds are successful in remembering three objects but failed at the task when the quantity was greater than three (Feigenson and Carey, 2003, 2005). Interestingly, 14-month-old infants can use high-level strategies such as chunking to remember more items (Feigenson and Halberda, 2004; Kibbe and Feigenson, 2016). Together, these findings demonstrate that the amount of information and the relationships between objects that can be maintained in VWM develops significantly between 6 and 14 months. However, many important questions remain open about how these processes operate in infants.

The influence of context on object perception has only recently been explored in infants. Examining eye gaze patterns of natural and artificial scenes, object-context congruency, and relational memory has revealed that 4-month-old infants fixate more on objects than the background in natural scenes (Bornstein et al., 2011a), and on objects that are congruent than incongruent with the scene context (Bornstein et al., 2011b). Nine-month-old infants can learn arbitrary face-scene associations (Richmond and Nelson, 2009), and by 12 months, some aspects of their scene scanning and fixation patterns are similar to adults' such as an early exploratory period with short fixations (Pannasch et al., 2008; Helo et al., 2016), and they also showing differences in the degree that saliency influenced eye movements (Helo et al., 2014).

Here we investigated how infants accumulate information to build a rich representation of objects embedded in scenes over time and interruptions, where interruptions consisted of an exposure to an intervening scene between repeat exposures of the same scene. Research in adults found that memory capacity estimates increased with exposure time when real-world objects were embedded in naturalistic scenes (Melcher, 2001, 2006; Brady et al., 2016). This is contrary to research using monochromatic, geometric objects without a rich background that report a plateau in performance after a certain exposure period (Luck and Vogel, 1997). The semantic richness of real-world stimuli and their familiarity was speculated to enhance memory performance (Brady et al., 2011). Increasing exposure times with these stimuli allowed adults to construct a more robust memory representation that was less prone to decay over time.

Another factor that can influence the robustness of visual memories is interruption that can disrupt the encoding and consolidation process. In the real world, objects are often occluded

for brief periods because of changes in the environment or changes in body positioning. Visual memory is essential in maintaining representations over these periods. However, some studies in adults found that brief interruptions caused by intervening stimuli did not significantly impact memory performance (Melcher, 2001, 2006; Melcher and Kowler, 2001). Surprisingly, adults demonstrated similar memory performance when presented with a continuous presentation of displays with objects in a scene compared to when the same displays were presented in intervals that added up to the same duration. That is, interruptions (even up to 20–30 s) did not interfere with the gradual accumulation of visual information. Together, these findings demonstrate both a gradual accumulation over time and persistence over brief periods of interference in adults for objects in scenes.

Only a few studies have explored the effects of interruptions in infant VWM encoding so far. Kaldy and Leslie (2005) reported that when 6-month-old infants saw two items hidden sequentially, they could only remember the features of the last-hidden object. A control study demonstrated that this failure was not due to decay over time: 6-month-olds were successful with the same occlusion time but without an interfering event (the hiding of the second object). In 10- to 14-month-olds, the maintenance of a memory trace was found to be dependent on the number of intervening items, and exceeding capacity limits lead to catastrophic forgetting (Feigenson and Carey, 2003, 2005; Cheries et al., 2006). These results indicate that although infants' memory capacity is increasing during the first year of life, their VWM is more susceptible to interference during maintenance and may be significantly less durable than adults'.

Our goal in the current study was to examine infants' ability to accumulate visual memories for objects in scenes and test whether those memories can persist over interruptions in order to identify factors contributing to infants' memory limitations in real-world settings. In adults, Melcher (2001, 2006, 2010) found increased accuracy with longer encoding periods with no decrement in performance when encoding was interrupted. We adapted this paradigm to be suitable for young infants. We manipulated exposure time to measure whether there was evidence of accumulation and introduced interruptions to investigate whether there is persistence of memories over repeated exposures. To evaluate infants' memory performance, we measured looking times to the changed object (a novelty preference-based process). We also tested a sample of adults to replicate the effects of accumulation and persistence using our stimuli and to serve as a comparison for infants' performance (with only minor procedural modifications, see section Materials and Methods). We hypothesized that similar to adults, longer encoding times will lead to improvements in infants' memory performance; however, unlike in adults, infants' performance will be lower when the same encoding time is broken up into multiple exposures.

## MATERIALS AND METHODS

In this experiment, infant and adult participants' visual memory was assessed using a change detection paradigm. Two experimental conditions were contrasted: continuous exposures and repeat

exposures (see **Table 1**). In continuous exposure trials, participants viewed a computer-generated scene with a fixed number of objects and exposure time was varied. This encoding phase was followed by a test display, where we measured whether participants could identify which of the objects had changed. In repeat exposure trials, an intervening scene was presented between the exposures of the scene. The final exposure was followed by a test display, just as in the continuous exposure condition. The two trial types were presented in a mixed block with trials presented in a fixed pseudorandom order. Manipulation of participants' encoding time allowed us to test memory accumulation and the manipulation in the number of repetitions of exposure the persistence of the memory for the objects in the scene.

## Participants

Twenty-three adults (female: 15, mean age =  $24.8 \pm 4.8$  years) participated in the adult version of the experiment. Adult participants were undergraduate and graduate students from the University of Massachusetts Boston. The participants were 56.5% Caucasian, 8.7% Black/African-American, and 34.8% Asian. The sample size was based on prior studies by Melcher ( $N = 6-21$ ; Melcher, 2001, 2006), and provides sufficient statistical power to detect differences between the conditions.

Nineteen full-term, healthy 12-month-old infants (female: 6, mean age =  $12;09 \pm 0;25$ , month; days) participated in the experiment. One infant was excluded from analysis due to fussiness, resulting in a final sample of 18 infants. Of the 75.7% of parents that provided information on racial background, 64.3% identified as Caucasian, 7.1% as Black/African-American, 14.3% as Asian, and the remaining 14.3% as multi-racial. We determined the appropriate sample size based on the results of pilot study with 8- to 16-month-old infants (Guillory et al., 2015), that showed that a minimum sample size of 11 infants was sufficient to detect a difference between conditions with 80% power and an alpha level of 0.05 (G\*Power 3.1, see Faul et al., 2007).

All participants had normal or corrected-to-normal vision with no history of colorblindness in the family. Written informed consent was obtained for all participants from the parent/legal guardian and the University of Massachusetts Boston's Institutional Review Board approved the study protocol.

## Apparatus and Stimuli

A Tobii T120 17-inch eye-tracker (Tobii Technology, Stockholm, Sweden) with a screen resolution of 1,024 pixels  $\times$  768 pixels at 32 bits per color and a refresh rate of 60 Hz was used for data collection. Eye gaze coordinates were collected at 60 Hz. The scenes (virtual rooms) were generated using the Sweet Home 3D software application. The rooms consisted of pieces of furniture, wallpaper, and floor tiling. There were 10 possible colors used for both the wallpaper and the furniture, and six

possible textures for the floor tiling. By manipulating the color combination of each feature, we generated 60 unique rooms (scenes). Furniture consisted of a collection of highly abstract cylindrical or block shapes, which created flat surface areas for the objects to be placed on, and for a given room consisting of approximately 4–6 surfaces. The objects were selected from a database (Blackleaf Studios, www.mygrafico.com) of colorful cartoon drawings of animals (**Figures 1, 2** provide examples of object placement). There were 40 different objects and each object was approximately of equal area, 17,450 pixels<sup>2</sup> (see examples in **Figure 1A**). Objects were placed within the scene to avoid any one object obstructing the other.

There were two versions of the task, one designed for infants that contained fewer objects in a trial, longer exposure times, and shorter inter-trial delays, henceforth referred to as the *infant version*, and a task designed for adults that had more objects, shorter exposure times, and longer inter-trial delay periods (*adult version*). Stimulus encoding times of 4 s in duration have demonstrated to be sufficient in achieving above-chance memory performance with 10-month-olds (Kaldy et al., 2016).

The *infant version* of the task contained three objects per scene (**Figure 1B**) that were placed on three of the 4–6 potential surfaces in the room. Tobii Studio 3.2 software was used to present and collect eye gaze data. Each object was shown once before the test trials during the object familiarization period. For the *adult version*, each room comprised of six objects that were placed on the top surface of the furniture that created the virtual scene and Psyscope XB70 software was used to present the stimuli and record manual responses. Individual differences in capacity limits have found a range of capacity estimates depending on parameters and test procedures; here we used six, as it seemed that that was unlikely to result in ceiling effects (Cowan, 2001; Brady et al., 2016).

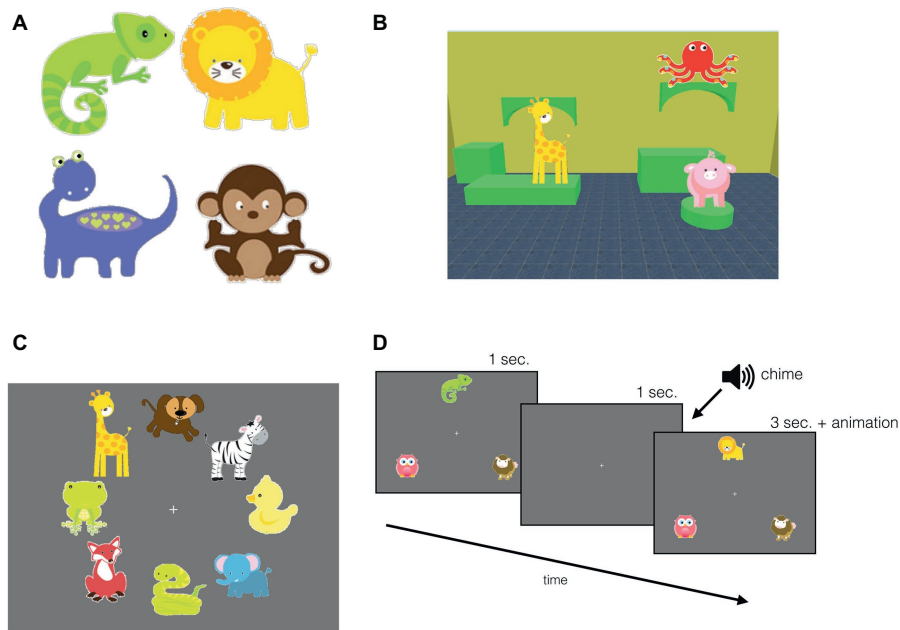
## Procedure

Infants viewed the videos while seated on a caregiver's lap, approximately 60 cm from the screen. The test session started with a standard infant-friendly 5-point calibration. The experiment consisted of three parts: object familiarization, task familiarization, and test trials that were run consecutively without a break, lasting approximately 5 min.

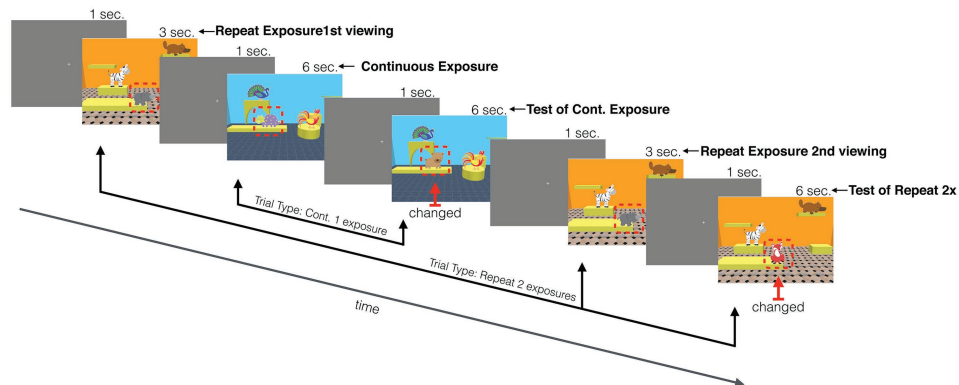
The first phase of the experiment began with object familiarization. To expose the participant to each object once and reduce novelty effects during the task, eight objects were displayed simultaneously on a grey background radially around the central fixation for 10 s (**Figure 1C**). A blank screen followed each display for 1 s (2 s in the *adult version*) with a fixation cross in the center. There were five object familiarization trials presented, exposing participants to all 40 objects that were used in the test trials. Next, following object familiarization, there were three task familiarization trials designed to expose participants to our change detection task: to familiarize them with the sequence of events and the chime which served as a signal that a test scene was about to appear, followed by the feedback animation. In the task familiarization trials, three objects appeared in a triangular formation with one object above central fixation and two to the bottom right and left

**TABLE 1 |** Trial types and durations (in seconds).

Group	Cont. (one exposure)	Two exposures	Four exposures
Adults	1, 2, 4	1 + 1, 2 + 2	1 + 1 + 1 + 1
Infants	3, 6	3 + 3	



**FIGURE 1 |** Experimental stimuli and familiarization trials. **(A)** A sample of the objects used in the study. **(B)** An example scene for the *infant version*: scene featuring three objects and the “room” comprised of a unique combination of furniture, wallpaper, and floor tiling. A sample display of **(C)** object familiarization, and **(D)** task familiarization trial sequence, where the top object changed identity after a 1-s delay.



**FIGURE 2 |** A schematic of the *infant version* of the task, showing a continuous trial interleaved between the repeat exposures (changed objects are shown with a red arrow and outlined with a dashed red line that was not present during actual testing).

of the central fixation on a grey background. The set of three objects were displayed for 1 s after which they disappeared. After 1 s, participants heard a chime intended to serve as a cue that a memory test for the previously presented display would follow. When the objects reappeared, one of the objects was replaced with a novel object. After 3 s, this novel object (target) was animated with an accompanying sound effect, which served as feedback (**Figure 1D**). These familiarization phases were incorporated in the experiment to diminish novelty effects in a similar fashion that recognition memory studies habituate infants to an image (Fagan, 1972, 1973). In the object familiarization trials, we presented all 40 objects at least once, while minimizing overall session duration. Although infants

might not have fixated all of the objects, the size of the attentional window can capture more than just the fixated objects, even in infants (Hernández et al., 2010; Ronconi et al., 2016). The goal of the task familiarization trials was to make learning the event sequence easier for infants: in these trials, the three objects always appeared in the same location to reduce the need to scan the display, and instead of a complex scene context, the objects were presented on a monochromatic background.

These two familiarization phases were followed immediately by the test trials. Each test trial started with a central fixation cross that was presented for 1 s (3 s in the *adult version*). Infants were then presented with a scene that was one of two duration lengths: 3 or 6 s (the *adult version* contained three

exposure lengths: 1, 2, and 4 s). A fixation screen followed this scene and then either a test scene (6 s for both *infants* and *adults*) or another scene was presented, depending on trial type: continuous vs. repeat exposures (**Figure 2** shows a schematic of a sequence of trials). A continuous exposure trial type consisted of one scene that was immediately followed by the test scene. The repeat exposure trial type consisted of an exposure to an initial scene, followed by an intervening, different scene trial, then a repeat exposure to the initial scene, and finally a test scene. This intervening scene involved different objects and room configuration. In the *infant version* with a maximum of two repeats, the intervening scene was always a continuous trial type, and in the *adult version* consisted of up to four repeats where the intervening scenes were of both continuous and repeat trials intermixed, similar to the procedure used in Melcher (2001, 2006). Which of the three locations had the changed object was randomly selected (the absolute locations were constrained by the room configuration of each scene).

Altogether, there were three different trial conditions in the *infant version* and six trial conditions in the *adult version* (**Table 1**). Adults were presented with 10 trials of each trial condition for a total of 60 trials. Infants were shown 15 trials: six trials were of the 3-s continuous exposure trial condition, five were the 3-s repeat exposure condition, and four trials were the 6-s continuous exposure condition. Before each memory test display, participants heard a chime. The test display consisted of three objects that appeared in the same locations as the objects shown during the exposure period with the exception that one of the three objects was replaced with a novel one. Adults were presented with test scenes where three of the original six objects were marked by numbers (that is, a partial report test). The test scene was always displayed for 6 s (in both the adult and the infant versions). For each age group (infant and adults), rooms, objects, and object placements were the same across participants.

## Data Analyses

### Adult Version

Following the procedures of Melcher (2001), adult participants were instructed to select *via* button press, one item out of the three marked objects that changed (selected from the set of six presented in the original scene). Each object in this marked subset was labeled as 1, 2, or 3 (the numbers appeared above the objects) and participants used a Dell keyboard to give their responses. Correct trials consisted of trials where the subject correctly identified the changed object within the 6 s of the test display (before the start of the feedback animation). Trials were categorized as incorrect when the participants selected an object that did not change during the response period (6 s) or responded after the end of this period (during the feedback animation).

### Infant Version

We calculated a preference measure based on proportional looking: during the test scene, the time spent looking at the

target object was divided by the total time spent looking at the three objects. This measure was compared to chance (33%). The default Tobii fixation filter was used for data analysis. Areas of interest were defined as equal-sized rectangular areas surrounding each of the objects (AOI size: 240 pixels × 240 pixels).

## Object and Task Familiarization

One-sample *t*-tests, Bonferroni corrected, were used to analyze the test phase of each of the three task familiarization trials. There were two missing values from two different infants that never looked at the screen during one of the three trials; therefore, instead of a repeated measures ANOVA, one-sample *t*-tests compared the proportion looking time to the target object in each of the three trials to chance performance (33%).

## Memory Accumulation

To analyze whether participants' memories accumulated over time, performance in the continuous trial types with the different durations were compared to each other. In the *adult version*, this analysis consisted of three different durations (1, 2, and 4 s) and in the *infant version*, two durations (3 and 6 s). Here, we applied a linear regression analysis to determine whether there was a relationship between the duration of exposure and memory performance. In addition, in the *infant version*, we performed one-sample *t*-tests comparing proportion looking results with chance performance (33%) to determine whether infants showed a novelty preference for the new object in the test scene in each trial type.

## Memory Persistence (Resistance to Interruption)

To analyze whether shorter repeated exposures were equivalent to a continuous exposure of the same total duration (e.g., a 2-s exposure repeated twice for a total duration of 4 s results in a similar memory performance in a continuous trial of 4 s), we performed a repeated measure analysis of variance (ANOVA) and paired sample *t*-tests on the proportion looking measure.

Additional exploratory analyses were performed to further examine infants' performance. We explored the persistence of memory between the two exposures in the repeat exposure trials. We compared proportion looking during the first vs. the second exposure per object and used one-sample *t*-tests to determine whether objects were viewed for similar lengths across exposures.

## Memory Accumulation (Infant vs. Adults)

Lastly, the regression coefficient,  $\beta$ , of adults and infants was compared, testing the null hypothesis that the coefficients are the same ( $\beta_{\text{adults}} = \beta_{\text{infants}}$ ). We achieved this by adding a predictor term to a regression model that reflected the interaction of the two factors [Group (adult/infant) and Encoding time] where the adult sample served as the reference group. The interaction term corresponded to the coefficient difference between groups ( $\beta_{\text{infants}} - \beta_{\text{adults}}$ ), such that no significant difference indicated no difference in slope.



## RESULTS

### Memory Accumulation—Adults

All trials were valid and included in the analysis. To determine whether there was a significant accumulation of information as encoding time increased, a linear regression was applied to predict performance from the total encoding time during the continuous trials (1, 2, and 4 s). Performance significantly increased with increased encoding time [ $F(1,67) = 11.85$ ,  $p = 0.001$ ], with a model fit of  $R^2 = 0.15$  (Figure 3). Participants' predicted accuracy increased by 3.8% for each second of additional encoding time,  $t(67) = 3.4$ ,  $p = 0.001$ . These results replicate prior findings in similar tasks that showed that increased encoding time improves recall performance for complex, real-world objects in scenes (Melcher, 2001; Brady et al., 2009b).

### Memory Persistence—Adults

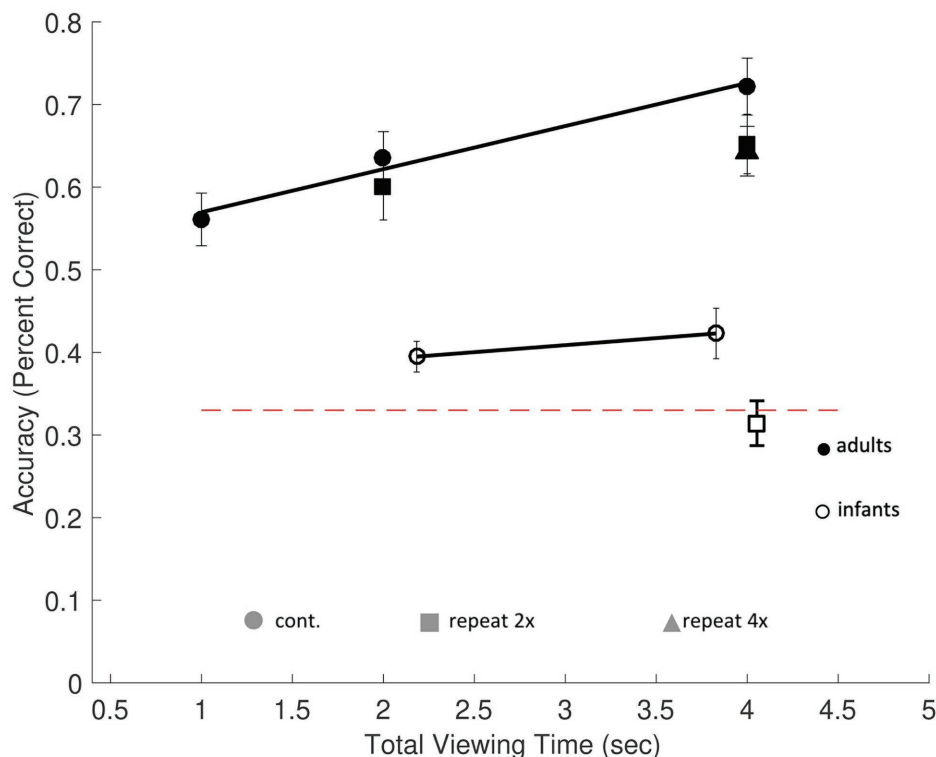
The persistence of the memory representations was tested with a repeated measures ANOVA by comparing the three conditions when the total exposure time was 4 s (4 s continuous, 2 s × 2 s repeat exposures, and 4 s × 1 s repeat exposures). The main effect of condition was not significant,  $F(2,44) = 1.82$ ,  $p = 0.17$ ,  $\eta_p^2 = 0.076$ . The same was true when comparing the two conditions where the total exposure time was 2 s (2 s continuous,

2 s × 1 s repeat exposures),  $F(1,22) = 0.67$ ,  $p = 0.42$ ,  $\eta_p^2 = 0.030$ . Together, these results suggest that viewing a scene for 4 continuous seconds is equivalent to viewing a scene twice for 2 s, or four times for 1 s, showing an essentially lossless memory representation in adults despite intervening scenes.

### Familiarization—Infants

Object familiarization trials: average looking time to individual objects during these trials was  $1.45 \pm 0.43$  s (means and standard errors are reported from here on). Results for each individual object within our set of 40 objects were within three SDs of this mean.

Task familiarization trials: applying one-sample  $t$ -test, the proportion looking at the target compared to chance (0.33) during the reappearance of the objects was significantly higher in the first trial [ $t(16) = 4.00$ ,  $p = 0.003$ ]. In the remaining two trials there was significant to a marginally significant difference of below chance looking [second:  $t(17) = -2.61$ ,  $p = 0.054$ ; third:  $t(16) = -2.82$ ,  $p = 0.036$ ]. Overall proportion looking was comparable across the three familiarization trials [the main effect of Trial was not significant:  $F(1.48,22.18) = 0.614$ ,  $p = 0.50$ ]. It should be noted that (unlike in the test trials), the objects in familiarization trials appeared in the same three canonic (left, right, top center) locations on each trial.



**FIGURE 3 |** Memory performance in Experiment 1. The rates of encoding time predicting memory performance (percent correct), in adults (filled symbols) and infants (open symbols), with chance performance defined as 33% (dashed red line). Accuracy was based on proportional looking time to the changed object during the test scene in infants and manual responses in adults. Average actual infant looking time on screen is represented on the horizontal axis. Results from continuous exposure trials (circles) and repeat exposure (square and triangle symbols) are shown. Memory accumulation effects among continuous exposure conditions are shown by the solid regression lines. Errors bars are  $\pm 1$  SEM.

## Test Performance Summary—Infants

Of the 15 test trials presented, on average  $3.4 \pm 0.7$  trials were excluded for each participant for insufficient eye gaze data. Valid trials required that infants looked at the scene during the encoding period ( $M = 2.51$  s,  $SD = 0.09$  s during exposure and  $M = 3.72$  s,  $SD = 0.21$  s during test). A minimum threshold for a fixation of at least (60 ms) within one of the three objects' AOIs was used, taking into account the limited temporal resolution of the Tobii T120 (Tatler and Vincent, 2008). Not surprisingly, infants typically did not look at the scenes continuously during the entire exposure period. Their actual average looking times were  $2.18 \pm 0.08$  s in the 3-s continuous exposure condition, a total of  $4.05 \pm 0.17$  s in the 3-s repeat exposure condition, and  $3.83 \pm 0.21$  s in the 6-s continuous exposure condition. In order to facilitate comparisons between infant and adult results, we used these observed looking time values when plotting infants' results (see **Figure 3**).

## Memory Accumulation—Infants

The continuous trial types (3 and 6 s) were analyzed to test for memory accumulation. One-sample *t*-tests were used to test whether infants' proportion of looking time at the changed object was different from chance. Infants performed significantly better than chance (33%) in both the 3-s continuous exposure [ $M = 0.39 \pm 0.02$ ,  $t(17) = 3.49$ ,  $p = 0.008$ ,  $d = 0.82$ ] and the 6-s continuous exposure condition [ $M = 0.42 \pm 0.03$ ,  $t(17) = 3.04$ ,  $p = 0.02$ ,  $d = 0.72$ ]. However, in the 3-s repeat condition, performance was not significantly different from chance, [ $M = 0.31 \pm 0.03$ ,  $t(17) = -0.58$ ,  $p = 1.00$ ,  $d = -0.12$ ].

In repeated exposure trials, for data to be included in these analyses, infants had to look at the scene during both exposures. Therefore, repeated exposure trials were less likely to meet our inclusion threshold than continuous exposure trials, and our final data set contained fewer valid trials in this condition ( $M = 3.22 \pm SE = 0.24$  vs.  $M = 3.39 \pm SE = 0.16$  in the 6-s and  $M = 5.00 \pm SE = 0.34$  in the 3-s continuous exposure conditions). We investigated whether a lower number of valid trials could have led to the higher variance found in the repeat exposure trials. Analyzing the proportion of valid trials for each condition in a repeated measures ANOVA, we found a significant main effect of trial type,  $F(2,34) = 21.24$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.56$ . *Post hoc* tests revealed that infants had a higher proportion of valid trials in the 3-s (proportion:  $M = 0.83 \pm 0.06$ ) and the 6-s ( $M = 0.85 \pm 0.04$ ) continuous conditions compared to the 2 s  $\times$  3 s repeat exposure ( $M = 0.64 \pm 0.05$ ) condition (Bonferroni-corrected,  $p < 0.001$ ) contributing to the higher variability of the results in this condition.

In an exploratory analysis, we relaxed our exclusion criteria and included those repeat exposure trials where infants only looked at the scene during the second exposure. When analyzing this expanded data set, there were no significant differences between conditions in the number of valid trials,  $F(1,4,23.77) = 0.24$ ,  $p = 0.71$ ,  $\eta_p^2 = 0.031$ . However, a one-sample *t*-test showed that despite a small increase in infants' overall performance ( $M = 0.33 \pm 0.03$ ), the overall pattern of results

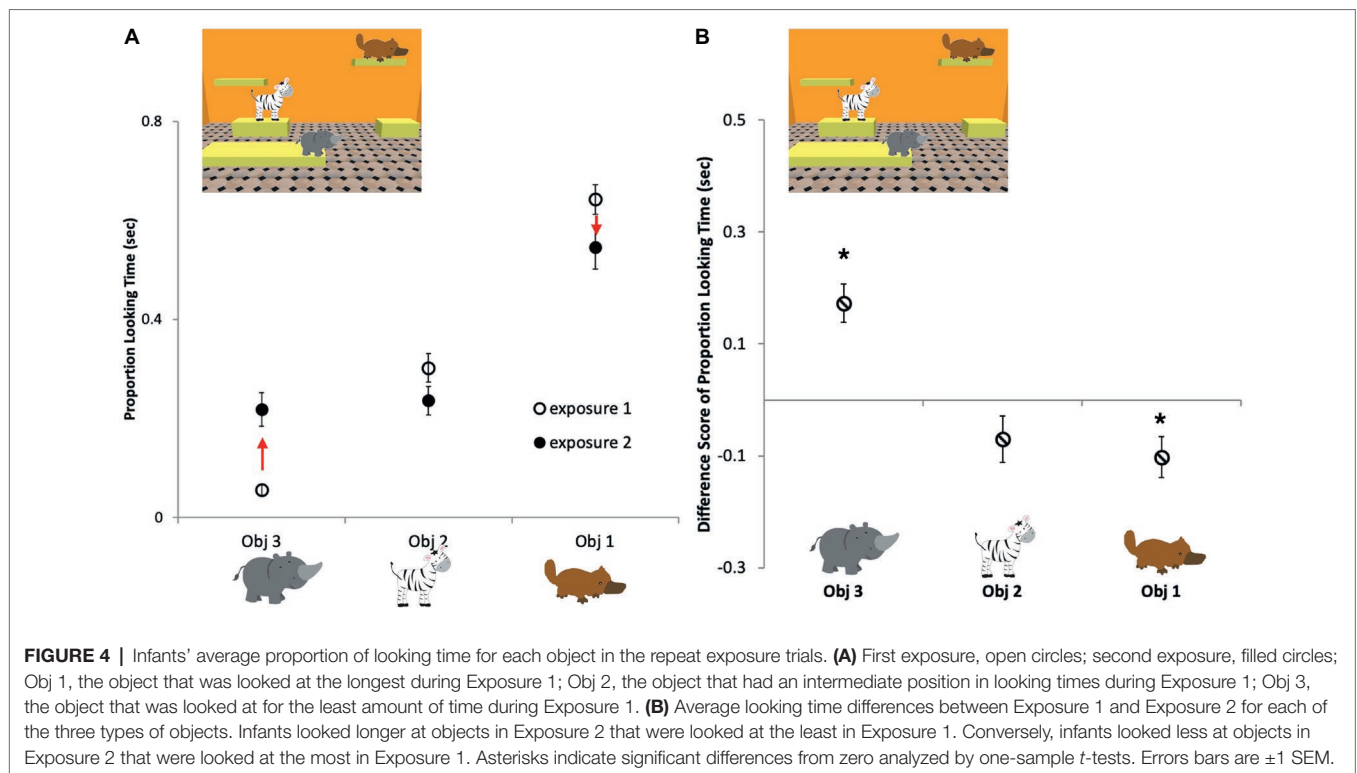
did not change, and it was still not significantly better than chance [ $t(17) = -0.06$ ,  $p = \text{n.s.}$ ]. Surprisingly, infants' performance was still lower in this expanded data set than in the 3-s continuous condition (0.33 vs. 0.39). Further studies are needed to clarify the source of this difference.

## Memory Persistence—Infants

To examine the persistence of memory representations in infants, we performed a paired sample *t*-test comparing performance in the two conditions when the total exposure time was equal, 6 s (6-s continuous vs. 2-s  $\times$  3-s repeat exposures). We found that memory performance significantly differed in the two conditions,  $t(17) = 2.66$ ,  $p = 0.017$ ,  $d = 0.63$ . Thus, we found that viewing a scene twice for 3 s was not equivalent to viewing a scene once for 6 s for infants.

Infants, unlike adults, cannot be instructed to use all of the available exposure time to encode the objects in the scenes; therefore, we analyzed the effects of infants' actual encoding times (their looking times during exposures) on memory performance. Results from the proportion looking during continuous exposure trials were subjected to a linear regression to test for effects of accumulation using average looking time for each subject in each condition (continuous trial types). Looking time during exposure was a marginally significant predictor of accuracy,  $F(1,34) = 3.80$ ,  $p = 0.059$ , with an overall model fit of  $R^2 = 0.101$ . Infants' predicted accuracy increased by 3.2% for each second of additional encoding time. That is, longer encoding times lead to a significant increase in infants' memory performance.

To further investigate whether there was any evidence for persistence in infants' memory over interruptions, in an exploratory analysis, we compared looking times to individual objects during the two exposures in the repeat exposure trials. The proportion of looking time at each of the three objects was calculated for the two exposures separately (**Figure 4A**) and then the proportion looking during Exposure 1 was subtracted from Exposure 2 to measure the change in the proportion looking at each object (**Figure 4B**). Using one-sample Bonferroni-corrected *t*-tests, we compared each object's proportion change value to zero, where zero represents that there was no change in looking time at the object between the first exposure (Exposure 1) of a repeat trial and the second exposure (Exposure 2) of the repeat trial. We found that objects that were looked at the longest initially in Exposure 1 were looked at for a smaller proportion of time during Exposure 2 [ $-0.10 \pm 0.04$  s,  $t(16) = -2.81$ ,  $p = 0.04$ ,  $d = -0.68$ ], objects scanned the least during Exposure 1 were looked at longer in Exposure 2 [ $0.16 \pm 0.03$  s,  $t(17) = 4.85$ ,  $p < 0.001$ ,  $d = -1.41$ ], while objects that were intermediately attended to, according to this measure during Exposure 1, were looked at approximately the same amount of time in Exposure 2 [ $-0.07 \pm 0.04$  s,  $t(16) = -1.69$ ,  $p = 0.32$ ,  $d = -0.41$ ]. These results indicate that infants had some recollection of the objects in the scene when they saw them for the second time and continued to explore them in a strategic way during the second exposure.



**FIGURE 4 |** Infants' average proportion of looking time for each object in the repeat exposure trials. **(A)** First exposure, open circles; second exposure, filled circles; Obj 1, the object that was looked at the longest during Exposure 1; Obj 2, the object that had an intermediate position in looking times during Exposure 1; Obj 3, the object that was looked at for the least amount of time during Exposure 1. **(B)** Average looking time differences between Exposure 1 and Exposure 2 for each of the three types of objects. Infants looked longer at objects in Exposure 2 that were looked at the least in Exposure 1. Conversely, infants looked less at objects in Exposure 2 that were looked at the most in Exposure 1. Asterisks indicate significant differences from zero analyzed by one-sample *t*-tests. Errors bars are  $\pm 1$  SEM.

## Memory Accumulation—Adults vs. Infants

Lastly, we tested whether infants have a slower rate of encoding in comparison to adults, comparing the regression coefficient ( $\beta$ ) of adults to infants. Interestingly, the accumulation rates of adults and infants did not differ significantly from one another ( $\beta_{\text{infants} - \text{adults}} = -0.021$ ,  $p = 0.43$ , **Figure 3**, solid black lines).

## DISCUSSION

The present study aimed to assess the mechanisms used to remember multiple objects in a quasi-naturalistic scene in 12-month-old infants. In particular, our goal was to measure two specific processes of visual memory: accumulation and persistence of visual information (Melcher, 2001; Brady et al., 2009b). We examined this question using a change detection task, contrasting continuous encoding periods with varying length and repeat exposures. While Melcher (2001) used verbal recall to assess memory performance in adults, in our version of the paradigm, we measured recognition memory to make the task appropriate for infants. (This modification did not affect the main pattern of results in adults, which replicated those found by Melcher).

In infants, eye-tracking was used to contrast looking time differences between changed and unchanged stimuli. We found that infants performed significantly better than chance in the continuous exposure conditions of both 3 and 6 s. Our findings are consistent with previous studies showing that 12-month-olds can succeed at VWM tasks involving three objects in a Violation-of-Expectation task using real-world, 3D objects (Kibbe and Leslie, 2013) and that 10-month-olds prefer a changing

stream with set-size 3 in a change detection task with a 250-ms delay (Ross-Sheehy et al., 2003).

We replicated previous findings of linear accumulation of visual information over exposure time in adults (Melcher, 2001, 2006). We used two approaches to assess accumulation in infants. First, we looked at percent correct responses in demonstrating a novelty preference for the changed item and found significant increases in performance when exposure time was increased. While our sample consisted of 12-month-olds, it is notable that another study that manipulated exposure times in younger infants found contrasting results. Kwon et al. (2014) found that doubling the exposure time from 500 to 1,000 ms did not have a significant impact on 6-month-olds' memory performance in a change detection task. Our study design differed from theirs in several ways: we used longer exposures, longer delays between exposure and test, the objects were embedded in scenes, and the infants we tested were older. Taken together, all these factors could have impacted why the infants in our study were able to construct a more durable memory representation with increased exposure.

Adults showed persistence in their memory representations in our study, demonstrating that interruptions did not significantly disrupt their encoding processes (replicating the findings of Melcher, 2001, 2006). We examined memory persistence (resistance to interference caused by intervening stimuli) in infants in two ways. First, we compared accuracy in the continuous (6 s) and repeat exposure (2 s  $\times$  3 s) conditions and we found significant differences between them. Infants' performance in the repeat exposure condition was not significantly better than chance. These results are puzzling as infants were successful

in the 3-s continuous exposure condition. In repeat exposure trials, even if infants forgot the objects that they have seen in the first exposure, the second exposure should have resulted in an above-chance outcome. Thus, we conducted an exploratory analysis comparing scanning patterns between the two exposures. This exploratory analysis showed that infants retained some memories between exposures, as they systematically continued their inspection of the objects that they did not look at first. These results are consistent with adult studies that demonstrate memory-guided attention whereby memories influence eye movements during visual exploration (Brockmole and Henderson, 2005; Ryals et al., 2015; Hutchinson et al., 2016).

Previous infant studies that have probed memory performance using a natural scene context have found, that like in adults, regularities that characterize natural scenes influence memory. Duh and Wang (2014) tested 15-month-old infants with objects placed in different natural scenes that were either congruent with the scene gist (fire hydrant in the grass) or not (yellow bottle in the grass). They found that infants often missed salient changes that preserved the overall scene gist, but when the scene gist was disrupted by a change in a non-salient object, infants were able to detect the change. Similarly, 24-month-old toddlers were shown to look longer at objects that were highly salient regardless of semantic consistency, that is, for both congruent and incongruent settings (Helo et al., 2017). In our study, all objects were equally congruent (or incongruent) with the scene gist, and similar to the toddlers in Helo et al. (2017), infants were successful at detecting an object change, indicating that infants were storing information about individual objects in the scene. While there is a lot known about how infants remember object/location pairings in paradigms when the objects are well-segmented and presented without a context, considerably less is known about context-dependent memory through which individual elements are integrated within a scene. Oakes et al. (2011) found that in the presence of spatial reference points (adding a grid around the to-be-remembered items), 6-month-olds' performance improved when just one object needed to be remembered, but not when the set size increased. Understanding failures and successes on these tasks requires a better understanding of infants' abilities to build robust associations about objects and their context.

The conclusions we can draw from our results have some limitations. These results likely underestimate infants' performance, as a very small portion of the images was reused (presented in two different test trials) in the infant study, conceivably leading to a certain amount of proactive interference.

Proactive interference arises when previously encoded information interferes with the current contents of working memory (Crowder, 1976), and it has been shown to affect VWM in adults (Makovski and Jiang, 2008). It is also conceivable that infants may show a mixture of preferences for familiar vs. novel objects (Aslin, 2007; Sivakumaran et al., 2018) in this particular paradigm.

In summary, the goal of the current study was to characterize the early development of two specific processes of visual memory for objects embedded in scenes. One of the main objectives was to test whether visual information accumulates over time in young infants. We established that infants performed significantly better than chance in detecting a change in one of three objects and we found memory benefits of increased encoding time on the target object. Our second objective was to investigate whether infants show persistence in information encoding over interruptions, and we found that while infants recognized the objects from previously shown scenes, this did not lead to better recognition performance at this age. Our results open up the field for future developmental work aimed at characterizing the processes underlying the buildup of visual memory representations of objects in scenes.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of The University of Massachusetts Boston's Institutional Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Massachusetts Boston's IRB. The legal guardian signed consent to participate in the research study after an experimenter reviewed the consent form and answered any questions that the guardian might have.

## AUTHOR CONTRIBUTIONS

SG was involved in the study design, data collection, analysis, and writeup. ZK supervised SG and was involved in study conception, design, analysis, and writeup.

## FUNDING

This research was supported in part by the National Institutes of Health Grant R15HD086658 awarded to ZK and the NIH Grant TL1TR001434.

## REFERENCES

- Alvarez, G. A., and Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. Sci.* 15, 106–111. doi: 10.1111/j.0963-7214.2004.01502006.x
- Aslin, R. N. (2007). What's in a look? *Dev. Sci.* 10, 48–53. doi: 10.1111/j.1467-7687.2007.00563.x
- Awah, E., Barton, B., and Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychol. Sci.* 18, 622–628. doi: 10.1111/j.1467-9280.2007.01949.x
- Bornstein, M. H., Mash, C., and Arterberry, M. E. (2011a). Young infants' eye movements over "natural" scenes and "experimental" scenes. *Infant Behav. Dev.* 34, 206–210. doi: 10.1016/j.infbeh.2010.12.010
- Bornstein, M. H., Mash, C., and Arterberry, M. E. (2011b). Perception of object-context relations: eye-movement analyses in infants and adults. *Dev. Psychol.* 47, 364–375. doi: 10.1037/a0021059
- Brady, T. F., Konkle, T., and Alvarez, G. A. (2009a). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *J. Exp. Psychol. Gen.* 138, 487. doi: 10.1037/a0016797



- Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vis.* 11:4. doi: 10.1167/11.5.4
- Brady, T. F., Konkle, T., Oliva, A., and Alvarez, G. A. (2009b). Detecting changes in real world objects: the relationship between visual long-term memory and change blindness. *Commun. Integr. Biol.* 2, 1–3. doi: 10.4161/cib.2.1.7297
- Brady, T. F., Störmer, V. S., and Alvarez, G. A. (2016). Working memory is not fixed-capacity: more active storage capacity for real-world objects than for simple stimuli. *Proc. Natl. Acad. Sci.* 113, 7459–7464. doi: 10.1073/pnas.1520027113
- Brockmole, J. R., and Henderson, J. M. (2005). Prioritization of new objects in real-world scenes: evidence from eye movements. *Hum. Percept. Perform.* 31, 857–868. doi: 10.1037/0096-1523.31.5.857
- Cherries, E. W., Wynn, K., and Scholl, B. J. (2006). Interrupting infants' persisting object representations: an object-based limit? *Dev. Sci.* 9, F50–F58. doi: 10.1111/j.1467-7687.2006.00521.x
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–185. doi: 10.1017/S0140525X01003922
- Cowan, N. (2010). The magical mystery four how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* 19, 51–57. doi: 10.1177/0963721409359277
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- Duh, S., and Wang, S. (2014). Infants detect changes in everyday scenes: the role of scene gist. *Cogn. Psychol.* 72, 142–161. doi: 10.1016/j.cogpsych.2014.03.001
- Fagan, J. F. III. (1972). Infants' recognition memory for faces. *J. Exp. Child Psychol.* 14, 453–476. doi: 10.1016/0022-0965(72)90065-3
- Fagan, J. F. III. (1973). Infants' delayed recognition memory and forgetting. *J. Exp. Child Psychol.* 16, 424–450. doi: 10.1016/0022-0965(73)90005-2
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146
- Feigenson, L., and Carey, S. (2003). Tracking individuals via object-files: evidence from infants' manual search. *Dev. Sci.* 6, 568–584. doi: 10.1111/1467-7687.00313
- Feigenson, L., and Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition* 97, 295–313. doi: 10.1016/j.cognition.2004.09.010
- Feigenson, L., and Halberda, J. (2004). Infants chunk object arrays into sets of individuals. *Cognition* 91, 173–190. doi: 10.1016/j.cognition.2003.09.003
- Fukuda, K., Vogel, E., Mayr, U., and Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychon. Bull. Rev.* 17, 673–679. doi: 10.3758/17.5.673
- Guillory, S. B., Gliga, T., and Kaldy, Z. (2018). Quantifying attentional effects on the fidelity and biases of visual working memory in young children. *J. Exp. Child Psychol.* 167, 146–161. doi: 10.1016/j.jecp.2017.10.005
- Guillory, S. B., Scipione, N., and Kaldy, Z. (2015). An eye-tracking study of infants' accumulation of visual memories in scenes. *Poster presented at the Biennial Meeting of the Society for Research in Child Development*, Philadelphia, PA.
- Helo, A., Pannasch, S., Sirri, L., and Rämä, P. (2014). The maturation of eye movement behavior: scene viewing characteristics in children and adults. *Vis. Res.* 103, 83–91. doi: 10.1016/j.visres.2014.08.006
- Helo, A., Rämä, P., Pannasch, S., and Meary, D. (2016). Eye movement patterns and visual attention during scene viewing in 3-to 12-month-olds. *Vis. Neurosci.* 33:E014. doi: 10.1017/S0952523816000110
- Helo, A., van Ommen, S., Pannasch, S., Danten-Dordoigne, L., and Rämä, P. (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behav. Dev.* 49, 248–266. doi: 10.1016/j.infbeh.2017.09.008
- Henderson, J. M., and Hollingworth, A. (1998). "Eye movements during scene viewing: an overview" in *Eye guidance in reading and scene perception*. ed. G. Underwood (Oxford, England: Elsevier), 269–293.
- Hernández, M., Costa, A., and Humphreys, G. W. (2010). The size of an attentional window affects working memory guidance. *Atten. Percept. Psychophys.* 72, 963–972. doi: 10.3758/APP.72.4.963
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 519–537. doi: 10.1037/0096-1523.30.3.519
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 31–47. doi: 10.1037/0096-1523.33.1.31
- Hollingworth, A., and Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Vis. Cogn.* 7, 213–235. doi: 10.1080/135062800394775
- Hutchinson, J. B., Pak, S. S., and Turk-Browne, N. B. (2016). Biased competition during long-term memory formation. *J. Cogn. Neurosci.* 28, 187–197. doi: 10.1162/jocn\_a\_00889
- Kaldy, Z., Guillory, S. B., and Blaser, E. (2016). Delayed match retrieval: a novel anticipation-based visual working memory paradigm. *Dev. Sci.* 19, 892–900. doi: 10.1111/desc.12335
- Kaldy, Z., and Leslie, A. M. (2003). Identification of objects in 9-month-old infants: integrating 'what' and 'where' information. *Dev. Sci.* 6, 360–373. doi: 10.1111/1467-7687.00290
- Kaldy, Z., and Leslie, A. M. (2005). A memory span of one? object identification in 6.5-month-old infants. *Cognition* 97, 153–177. doi: 10.1016/j.cognition.2004.09.009
- Kibbe, M. M., and Feigenson, L. (2016). Infants use temporal regularities to chunk objects in memory. *Cognition* 146, 251–263. doi: 10.1016/j.cognition.2015.09.022
- Kibbe, M. M., and Leslie, A. M. (2011). What do infants remember when they forget? Location and identity in 6-month-olds' memory for objects. *Psychol. Sci.* 22, 1500–1505. doi: 10.1177/0956797611420165
- Kibbe, M. M., and Leslie, A. M. (2013). What's the object of object working memory in infancy? Unraveling 'what' and 'how many'. *Cogn. Psychol.* 66, 380–404. doi: 10.1016/j.cogpsych.2013.05.001
- Kwon, M.-K., Luck, S. J., and Oakes, L. M. (2014). Visual short-term memory for complex objects in 6- and 8-month-old infants. *Child Dev.* 85, 564–577. doi: 10.1111/cdev.12161
- Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846
- Makovski, T., and Jiang, Y. (2008). Proactive interference from items previously stored in visual working memory. *Mem. Cogn.* 26, 43–52. doi: 10.3758/MC.36.1.43
- Melcher, D. (2001). Persistence of visual memory for scenes. *Nature* 412:401. doi: 10.1038/35086646
- Melcher, D. (2006). Accumulation and persistence of memory for natural scenes. *J. Vis.* 6, 8–17. doi: 10.1167/6.1.2
- Melcher, D. (2010). Accumulating and remembering the details of neutral and emotional natural scenes. *Perception* 39, 1011–1025. doi: 10.1068/p6670
- Melcher, D., and Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vis. Res.* 41, 3597–3611. doi: 10.1016/S0042-6989(01)00203-6
- Oakes, L. M., Baumgartner, H. A., Kanjlia, S., and Luck, S. J. (2017). An eye tracking investigation of color-location binding in infants' visual short-term memory. *Infancy* 22, 584–607. doi: 10.1111/inf.12184
- Oakes, L. M., Hurley, K. B., Ross-Sheehy, S., and Luck, S. J. (2011). Developmental changes in infants' visual short-term memory for location. *Cognition* 118, 293–305. doi: 10.1016/j.cognition.2010.11.007
- Oakes, L. M., Ross-Sheehy, S., and Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychol. Sci.* 17, 781–787. doi: 10.1111/j.1467-9280.2006.01782.x
- Pannasch, S., Helmert, J. R., Roth, K., Herbold, A. K., and Walter, H. (2008). Visual fixation durations and saccade amplitudes: shifting relationship in a variety of conditions. *J. Eye Mov. Res.* 2, 1–19. doi: 10.16910/jemr.2.2.4
- Richmond, J., and Nelson, C. A. (2009). Relational memory during infancy: evidence from eye tracking. *Dev. Sci.* 12, 549–556. doi: 10.1111/j.1467-7687.2009.00795.x
- Ronconi, L., Franchin, L., Valenza, E., Gori, S., and Facoetti, A. (2016). The attentional 'zoom-lens' in 8-month-old infants. *Dev. Sci.* 19, 145–154. doi: 10.1111/desc.12288
- Rose, S. A., Feldman, J. F., and Jankowski, J. J. (2001). Visual short-term memory in the first year of life: capacity and recency effects. *Dev. Psychol.* 37, 539–549. doi: 10.1037/0012-1649.37.4.539
- Ross-Sheehy, S., Oakes, L. M., and Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Dev.* 74, 1807–1822. doi: 10.1046/j.1467-8624.2003.00639.x
- Ryals, A. J., Wang, J. X., Polnaszek, K. L., and Voss, J. L. (2015). Hippocampal contribution to implicit configuration memory expressed via eye movements during scene exploration. *Hippocampus* 25, 1028–1041. doi: 10.1002/hipo.22425

- Scholl, B. J., and Xu, Y. (2001). The magical number 4 in vision. *Behav. Brain Sci.* 24, 145–146. doi: 10.1017/S0140525X01523924
- Simmering, V. R. (2012). The development of visual working memory capacity during early childhood. *J. Exp. Child Psychol.* 111, 695–707. doi: 10.1016/j.jecp.2011.10.007
- Sivakumaran, M. H., Mackenzie, A. K., Callan, I. R., Ainge, J. A., and O'Connor, A. R. (2018). The discrimination ratio derived from novel object recognition tasks as a measure of recognition memory sensitivity, not bias. *Sci. Rep.* 8:11579. doi: 10.1038/s41598-018-30030-7
- Tatler, B. W., and Vincent, B. T. (2008). Systematic tendencies in scene viewing. *J. Eye Mov. Res.* 2, 1–18. doi: 10.16910/jemr.2.2.5
- Torralba, A., Oliva, A., Castelano, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766–786. doi: 10.1037/0033-295X.113.4.766
- Vogel, E. K., Woodman, G. F., and Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 92–114. doi: 10.1037//0096-1523.27.1.92
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2019 Guillory and Kaldy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Effect of Emotional Valence and Arousal on Visuo-Spatial Working Memory: Incidental Emotional Learning and Memory for Object-Location

Marco Costanzi<sup>1\*</sup>, Beatrice Cianfanelli<sup>1</sup>, Daniele Saraulli<sup>1</sup>, Stefano Lasaponara<sup>1,2</sup>, Fabrizio Doricchi<sup>2,3</sup>, Vincenzo Cestari<sup>2</sup> and Clelia Rossi-Arnaud<sup>2\*</sup>

<sup>1</sup> Department of Human Sciences, LUMSA University, Rome, Italy, <sup>2</sup> Department of Psychology, Sapienza University, Rome, Italy, <sup>3</sup> Fondazione Santa Lucia IRCCS, Rome, Italy

## OPEN ACCESS

### Edited by:

Zaifeng Gao,  
Zhejiang University, China

### Reviewed by:

Yixuan Ku,  
East China Normal University, China  
Raoul Bell,  
Heinrich Heine University  
of Düsseldorf, Germany

### \*Correspondence:

Marco Costanzi  
m.costanzi@lumsa.it  
Clelia Rossi-Arnaud  
clelia.rossi-arnaud@uniroma1.it

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 26 July 2019

**Accepted:** 31 October 2019

**Published:** 19 November 2019

### Citation:

Costanzi M, Cianfanelli B,  
Saraulli D, Lasaponara S, Doricchi F,  
Cestari V and Rossi-Arnaud C (2019)  
The Effect of Emotional Valence  
and Arousal on Visuo-Spatial Working  
Memory: Incidental Emotional  
Learning and Memory  
for Object-Location.  
Front. Psychol. 10:2587.  
doi: 10.3389/fpsyg.2019.02587

Remembering places in which emotional events occur is essential for individual's survival. However, the mechanisms through which emotions modulate information processing in working memory, especially in the visuo-spatial domain, is little understood and controversial. The present research was aimed at investigating the effect of incidentally learned emotional stimuli on visuo-spatial working memory (VSWM) performance by using a modified version of the object-location task. Eight black rectangles appeared simultaneously on a computer screen; this was immediately followed by the sequential presentation of eight pictures (selected from IAPS) superimposed onto each rectangle. Pictures were selected considering the two main dimensions of emotions: *valence* and *arousal*. Immediately after presentation, participants had to relocate the rectangles in the original position as accurately as possible. In the first experiment arousal and valence were manipulated either as between-subject (Experiment 1A) or as within-subject factors (Experiment 1B and 1C). Results showed that negative pictures enhanced memory for object location only when they were presented with neutral ones within the same encoding trial. This enhancing effect of emotion on memory for object location was replicated also with positive pictures. In Experiment 2 the arousal level of negative pictures was manipulated between-subjects (high vs. low) while maintaining valence as a within-subject factor (negative vs. neutral). Objects associated with negative pictures were better relocated, independently of arousal. In Experiment 3 the role of emotional valence was further ascertained by manipulating valence as a within-subject factor (neutral vs. negative in Experiment 3A; neutral vs. positive in Experiment 3B) and maintaining similar levels of arousal among pictures. A significant effect of valence on memory for location was observed in both experiments. Finally, in Experiment 4, when positive and negative pictures were encoded in the same trial, no significant effect of valence on memory for

object location was observed. Taken together results suggest that emotions enhance spatial memory performance when neutral and emotional stimuli compete with one another for access into the working memory system. In this competitive mechanism, an interplay between valence and arousal seems to be at work.

**Keywords:** working memory, emotional valence, arousal, object relocation, incidental encoding

## INTRODUCTION

In our daily lives, we experience and remember many features of an event that triggers an emotional response. For instance, we well remember places where emotional experiences took place. Thus, emotional stimuli and spatial information are encoded, linked, and stored in memory for future recall.

The mechanisms through which emotions modulate long-term memory consolidation have been widely investigated (McGaugh, 2000, 2004, 2006; Phelps, 2004; Richter-Levin, 2004; Anderson et al., 2006a,b; Cestari et al., 2006; La Bar, 2007; Mather, 2007; Walker, 2010; McReynolds and McIntyre, 2012; Leventon et al., 2018). Emotional stimuli are usually classified by considering two main dimensions: *Valence*, which describes the attractiveness (positive valence) or aversiveness (negative valence) of stimuli along a continuum (negative – neutral – positive), and *arousal*, which refers to the perceived intensity of an event from very calming to highly exciting or agitating (Kensinger and Schacter, 2006). It is well established, for long-term memory, that positive and negative arousing stimuli are better remembered than neutral non-arousing ones (Canli et al., 2000; Dolcos and Cabeza, 2002; Dolcos et al., 2004; Kensinger et al., 2006; Kensinger and Schacter, 2007; Kensinger et al., 2007; La Bar, 2007; Kensinger, 2009). Moreover, long-term memory is enhanced also for neutral stimuli by increasing participants' arousal, either through the presentation of emotional pictures (Anderson et al., 2006a; Ventura-Bort et al., 2016) or through the administration of chemical compounds usually released during emotional arousal (Buchanan and Lovall, 2001). The latter suggests the existence of an emotional tagging which is able to increase the salience of non-emotional stimuli (Richter-Levin and Akirav, 2003).

Despite the large amount of evidence on the effect of emotion on long-term memory, the mechanisms through which emotion influence working memory are not completely understood. Contrasting results have been reported, particularly as far as the visuospatial domain is considered (Kensinger and Corkin, 2003; Shackman et al., 2006; Mather, 2007; Levens and Phelps, 2008; Mather and Nesmith, 2008; Lindstrom and Bohlin, 2011; Mather and Sutherland, 2011; Bannerman et al., 2012; Gonzalez-Garrido et al., 2015; Tavares et al., 2016).

As in long-term memory studies, the role of emotion on working memory can be investigated either by modulating the emotional content of the stimuli used in the memory task or by manipulating the mood of participants.

Manipulations of participants' mood are usually aimed at enhancing the arousal level before performing working memory tasks. These manipulations include the presentation of affective video clips, the administration of stress-related compounds, or

the exposure to threatening conditions. Overall, studies on the effect of emotional arousal on working memory performance have produced mixed results. Further, interpreting results of experiments which have manipulated emotional arousal and mood is not always straightforward. In particular, problems of internal validity (i.e., difficulties in monitoring whether the adopted manipulation produced the desired effect on arousal or mood), difficulties in monitoring whether the effects of manipulation lasted throughout the memory task, and difficulties in determining on which phase of the task the manipulation is effectively acting (e.g., encoding or retrieval) have been reported (see Mather, 2007; Moran, 2016 for reviews).

Another procedure that has been effectively used to investigate how emotional information is processed in working memory consists in manipulating the emotional content of the stimuli used in the memory task (Mather, 2007; Pessoa, 2008, 2009; Kensinger, 2009; Mather and Sutherland, 2011). Using a modified version of the Corsi-block task, in which spatial positions were signaled by emotional or non-emotional stimuli (schematic faces, real faces or pictures), Bannerman et al. (2012) found that emotions did not affect spatial working memory performance, although emotional stimuli were able to capture attention more effectively than neutral ones (Bannerman et al., 2012). Differently, in a study in which participants had to remember spatial sequences of six facial (happy and neutral) and non-facial stimuli in reverse order, a performance improvement was observed in trials in which happy faces appeared (Gonzalez-Garrido et al., 2015). However, the type of stimuli used in the latter study (faces), the length of the task and the repeated presentation of the stimuli make it difficult to rule out the involvement of a competition mechanism between emotional and non-emotional stimuli for access to working memory (an aspect which we will further discuss later on).

In another study, a negative impact of emotions on working memory performance was observed (Yoon et al., 2016). The latter used a visual working memory test and participants were instructed to maintain (forward trials) or reverse (backward trials) the order of four emotional or four neutral pictures. Furthermore, a distraction effect exerted by emotional stimuli was observed in a digit recognition task with a high working memory load, suggesting that memory reduction could be induced by a depletion of attentional resources exerted by the emotional material (Tavares et al., 2016).

In the above-mentioned studies, emotional and non-emotional stimuli were presented in separate trials of the task and conclusions regarding the effect of emotion on working memory performance were drawn by comparing the performance achieved in trials in which only emotionally valenced stimuli were presented with the performance achieved



in trials in which only neutral stimuli were presented. Therefore, if one considers the emotional valence of stimuli administered during the working memory task, participants processed only one type (negative, positive, or neutral) of stimulus at a time.

On the other hand, when emotional and neutral stimuli are presented within the same encoding trial, a positive impact of emotion on memory performance always emerges. In an incidental encoding task in which both emotionally arousing and non-arousing pictures, selected from the International Affective Picture System (IAPS), appeared in different locations of the screen within the same encoding trial, participants better recognized the position of emotionally arousing pictures than the position of non-arousing ones. Interestingly, the performance improvement for emotionally arousing pictures was independent of their valence (negative or positive) (Mather and Nesmith, 2008). Therefore, the authors suggested that emotionally arousing stimuli had a priority access to the working memory system (Mather and Nesmith, 2008; Mather and Sutherland, 2011). In the same vein, Schmidt et al. (2011), by using a spatial and temporal recognition task in which mixed lists of both arousing and non-arousing emotional pictures were presented, found that memory for spatial location and temporal order for high-arousing stimuli was increased regardless of their valence (Schmidt et al., 2011). More recently, Schumann et al. (2018) found that memory recognition for emotional pictures was higher than memory for neutral pictures immediately after encoding a mixed list of positive, negative and neutral pictures (Schumann et al., 2018). Overall, the latter results are consistent with the Arousal-biased competition (ABC) theory, which predicts that encoding of within-object characteristics, like spatial location, is enhanced for arousing stimuli (Mather and Sutherland, 2011).

However, aside from arousal levels, the hypothesis that emotional valence is also important in modulating working memory performance has recently been discussed (Adelman and Estes, 2013; Kang et al., 2014; Wilson et al., 2016). In particular, an effect of valence on visuo-spatial working memory (VSWM) performance has been suggested in a recognition task in which memory for dot locations was increased by presenting negative pictures on the same side of the screen in which dots appeared (Wilson et al., 2016).

Putman et al. (2004) investigated the effect of valence on spatial working and long-term memory in healthy young women by using a face relocation task, in which both emotional (happy and fearful) and non-emotional faces were simultaneously encoded. Happy faces were better relocated than neutral ones in the immediate test, while both negative and happy faces were better relocated than neutral ones in the long-term memory test (Putman et al., 2004). These results suggest a valence-specific effect for the immediate memory performance, but not for the memory performance at long-term (Putman et al., 2004).

Considering the literature reported above, it is possible to hypothesize that emotions increase spatial memory when neutral and emotional stimuli are presented within the same encoding trial, suggesting the existence of a competition mechanism. Further, the role played by arousal and valence in regulating the

competitive access to the working memory system is not clear. These two aspects deserve investigation.

In the present study, we sought (a) to verify the competition hypothesis and (b) to ascertain the effect of valence and arousal in determining access to VSWM. To this purpose, the effect of incidentally learned emotional stimuli on VSWM performance was investigated. We used a modified version of the object-relocation task in which participants had to encode and remember the position of eight black rectangles simultaneously presented. While the rectangles were on the screen, emotional and non-emotional pictures were presented sequentially and overlapped for a brief period of time the rectangles. The ability to relocate the objects was tested immediately after presentation while the incidental learning of the emotional or non-emotional pictures was evaluated 24 h after the visuo-spatial memory test, using a free recall test. In Experiment 1A, the pictures associated to the rectangles during encoding were either all negative or all neutral (no competition), while in Experiment 1B, half of the rectangles were accompanied by negative pictures while the other half by neutral pictures (competition). Following the competition hypothesis, we expected an effect of emotion only in the second experiment. Similarly, Experiment 1C was performed by associating neutral and positive pictures (the latter of the same arousal level as negative pictures used in the previous experiment) to the rectangles, in order to explore competition mechanisms also with positive valenced stimuli. Further experiments were performed to better clarify which emotional feature affects spatial working memory performance in this paradigm. More specifically, in Experiment 2 we examined the effect of arousal by comparing the level of performance, under competition conditions i.e., with both negative and neutral pictures being presented in the same trial, with either low or high arousal pictures. Since in all previous experiments emotional and neutral pictures used within the same trial to elicit competition differed in both valence and arousal, we performed a further experiment in which we manipulated the valence of pictures, keeping the arousal level constant. To this purpose, in Experiment 3A we presented negative and neutral pictures with similar low levels of arousal, and in Experiment 3B positive and neutral pictures again with similar low levels of arousal. Since all the experiments performed to examine competition entailed the presentation of emotional versus neutral stimuli within the same trial, we designed a last experiment in order to verify the competition hypothesis when emotional stimuli with different valence (positive and negative), but similar arousal levels, compete with one another within the same trial (Experiment 4).

## MATERIALS AND METHODS

### Participants

A total of 226 (154 females; age:  $23.24 \pm 4.19$ ) undergraduate students in Psychology at LUMSA and Sapienza Universities voluntarily participated in the experiments. This study was carried out in accordance with the recommendations of Committees for Ethics, Department of Psychology Sapienza,

University of Rome, and LUMSA University. All subjects gave written informed consent in accordance with the Declaration of Helsinki. All participants were Italian speakers and reported having normal or corrected-to-normal vision.

## Materials and Procedures

To study the effect of the incidental presentation of emotional stimuli on VSWM performance, a modified version of the object relocation task (Kessels et al., 1999) was administered (see **Figure 1**). Participants sat in front of a PC screen and an instruction slide was shown in which they received information about the object-relocation task, but not about the presentation of emotional pictures.

Thirty-six pictures were selected from the International Affective Picture System (IAPS; Lang et al., 2008). Mean valence and arousal values for the pictures used in each experiment are reported in **Table 1**, and IAPS codes in the **Appendix**.

When participants felt ready, they pressed a button to begin the encoding phase which was signaled by the presentation of a cross in the center of the screen for 1000 ms. Immediately after cross presentation, eight black rectangles ( $170 \times 128$  px; 72 dpi) simultaneously appeared in random positions of the screen. After 1000 ms, eight pictures selected from IAPS appeared one at a time superimposed on each rectangle. Each picture was presented for 1000 ms (ISI: 250 ms). Thus, the encoding phase lasted 11 s and 750 ms. The test phase took place immediately after the end of the encoding phase. All the black rectangles appeared at the bottom of the screen and participants had to relocate them as accurately as possible, using the mouse. Memory for object location was evaluated considering the distance between the original position and the closest relocated object. Long-term

memory for incidentally learned pictures was evaluated 24 h later by a free recall task: participants were asked to verbally recall, by speaking the name of the objects depicted in the pictures they had seen on the previous day. After the free recall test, pictures were again presented one at a time, for 7 s, on the screen. Participants were instructed to view each picture and to subjectively evaluate valence and arousal of each picture by using the Self-Assessment Manikin (SAM).

## Experiment 1

### Experiment 1A

In Experiment 1A object positions were tagged by presenting negative pictures to one group (Negative group) and neutral pictures to a second group (Neutral group), in a between-subject manipulation. Therefore, the effect of emotional learning on VSWM was investigated in this first experiment in a “non-competitive” fashion. In brief, participants (32 females and 8 males; age:  $23.32 \pm 4.25$ ) were randomly assigned to two different groups: (i) *Negative* group in which object positions were tagged by negative IAPS pictures; (ii) *Neutral* group in which object positions were tagged by neutral IAPS pictures.

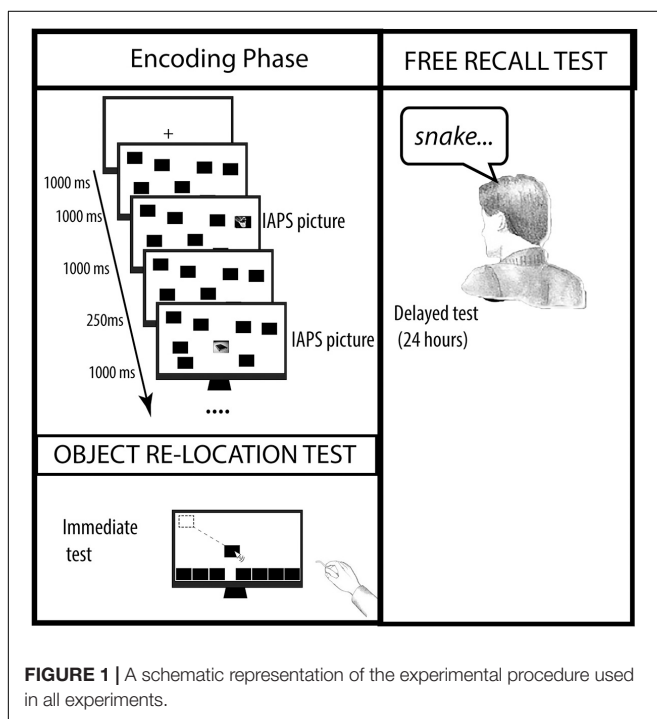
Since in this experiment a between- subject design was used, to further investigate if performance would be affected by basal differences in motor and spatial abilities of participants we performed a supplemental experiment in which a control condition followed the experimental trial with IAPS picture. In this control condition, the procedure was identical to that previously described but the images associated to rectangles were built by scrambling pixels of different colors.

### Experiment 1B

In Experiment 1B, within a single trial, half of the object positions were tagged by the presentation of neutral pictures and the other half by negative pictures, thus allowing a within-subject manipulation. Therefore, the effect of emotional learning on VSWM was investigated in a “competitive” fashion. In brief, after participants (20 females and 8 males; age:  $23.18 \pm 2.43$ ) simultaneously watched all eight objects for 1000 ms, four negative and four neutral pictures appeared superimposed on each object. The starting position of picture presentation was varied and each of the eight rectangles on the display could act as the starting position, thus yielding eight different configurations. Further, for half of the participants the first item presented was an emotional picture (negative in Experiments 2, 3, and 5; positive in Experiment 4), followed by a neutral one, while for the other half, presentation started with a neutral picture followed by an emotional one, resulting in 16 different configurations. The order was counterbalanced across participants.

### Experiment 1C

The third experiment was aimed at investigating the effect of emotional learning on VSWM in a “competitive” fashion, like in the second experiment, but object positions were tagged by the presentation of either positive or neutral pictures within a single trial. In brief, after participants (19 females and 6 males; age:  $24.4 \pm 3.58$ ) simultaneously watched all eight objects for 1000 ms, four positive and four neutral pictures appeared superimposed on



**TABLE 1** | Valence and arousal values of IAPS pictures used in the experiments.

	Pictures (n)	Valence (mean $\pm$ SD)		Arousal (mean $\pm$ SD)	
Exp. 1A	Negative (8) Neutral (8)	3.12 $\pm$ 0.26 4.69 $\pm$ 0.43	$t(14) = -8.912$ ; $p < 0.0001$	5.7 $\pm$ 0.17 2.6 $\pm$ 0.25	$t(14) = 28.496$ ; $p < 0.0001$
Exp. 1B	Negative (4) Neutral (4)	3.10 $\pm$ 0.26 4.69 $\pm$ 0.25	$t(6) = -8.85$ ; $p < 0.0001$	5.7 $\pm$ 0.25 2.5 $\pm$ 0.17	$t(6) = 21.34$ ; $p < 0.0001$
Exp. 1C	Positive (4) Neutral (4)	7.73 $\pm$ 0.31 4.69 $\pm$ 0.25	$t(6) = 9.12$ ; $p < 0.0001$	5.76 $\pm$ 0.37 2.5 $\pm$ 0.17	$t(6) = 16.03$ ; $p < 0.0001$
Exp. 2	Negative – High arousal (4) Negative – Low arousal (4) Neutral – High arousal (4) Neutral – Low arousal (4)	2.55 $\pm$ 0.77 3.1 $\pm$ 0.26 4.91 $\pm$ 0.04 4.7 $\pm$ 0.25	$F(3,12) = 30.42$ ; $p < 0.0001$ $p < 0.0001$ Neg vs. Neutral	7.12 $\pm$ 0.19 5.7 $\pm$ 0.25 2.02 $\pm$ 0.32 2.47 $\pm$ 0.17	$F(2,12) = 4.13$ ; $p < 0.0001$ $P < 0.0001$ Neg-Low vs. Neg-High $P < 0.0001$ Neg vs. Neutral
Exp. 3A	Negative (4) Neutral (4)	3.12 $\pm$ 0.72 4.72 $\pm$ 0.33	$t(6) = 4.02$ ; $p < 0.01$	3.81 $\pm$ 0.21 3.72 $\pm$ 0.18	$t(6) = 0.64$ ; $p = \text{n.s.}$
Exp. 3B	Positive (4) Neutral (4)	7.85 $\pm$ 0.15 4.72 $\pm$ 0.33	$t(6) = -17.13$ ; $p < 0.01$	3.81 $\pm$ 0.54 3.72 $\pm$ 0.18	$t(6) = -2.28$ ; $p = \text{n.s.}$
Exp. 4	Negative (4) Positive (4)	3.10 $\pm$ 0.26 7.96 $\pm$ 0.39	$t(6) = -20.93$ ; $p < 0.0001$	5.7 $\pm$ 0.25 5.68 $\pm$ 0.33	$t(6) = 0.09$ ; $p = \text{n.s.}$

Range of all IAPS pictures ( $n = 1194$ ) valence from 1.31 to 8.34; arousal from 1.72 to 7.35. Range IAPS pictures used in the present experiments ( $n = 36$ ) valence from 1.62 to 8.34; arousal from 1.72 to 7.35.

each object. The order of picture presentation was randomized like in Experiment 2.

## Experiment 2

The second experiment was aimed at investigating whether the increase in picture's arousal enhanced VSWM performance. Participants (30 females and 19 males; age:  $23.64 \pm 4.29$ ) were submitted to the object relocation task following the same procedure as in the previous experiment 1B. They were randomly assigned to two different groups: (i) *High arousal* in which object positions were tagged by four negative pictures with high arousal values, and four neutral pictures; (ii) *Low arousal* in which object positions were tagged by four negative pictures with low arousal values, and four neutral pictures.

## Experiment 3

### Experiment 3A

The present experiment was designed to investigate the effect of emotional valence on VSWM performance. Participants (11 females and 9 males; age:  $21.2 \pm 2.85$ ) were submitted to the object relocation task following the same procedure as in the Experiment 1B, but object positions were tagged by the presentation of four negative and four neutral pictures with similar arousal values.

### Experiment 3B

Participants (17 females and 4 males; age:  $23.4 \pm 2.79$ ) were submitted to the object relocation task following the same procedure as previously, but object positions were tagged by the presentation of four positive and four neutral pictures with similar arousal values.

## Experiment 4

The last experiment was designed to further investigate the effect of emotional valence on VSWM performance. Eighteen

participants (13 females; age:  $23.6 \pm 2.15$ ) were submitted to the object relocation task. The procedure was identical to the one used previously but the pictures differed in order to have four negative and four neutral pictures with similar levels of arousal.

## Data Analysis

In all experiments the displacement error was calculated as the distance (expressed in pixel) between the center of the originally positioned object and the center of the closest relocated object. For the evaluation of picture's memory, the proportion of correctly recalled pictures was considered for statistical analyses.

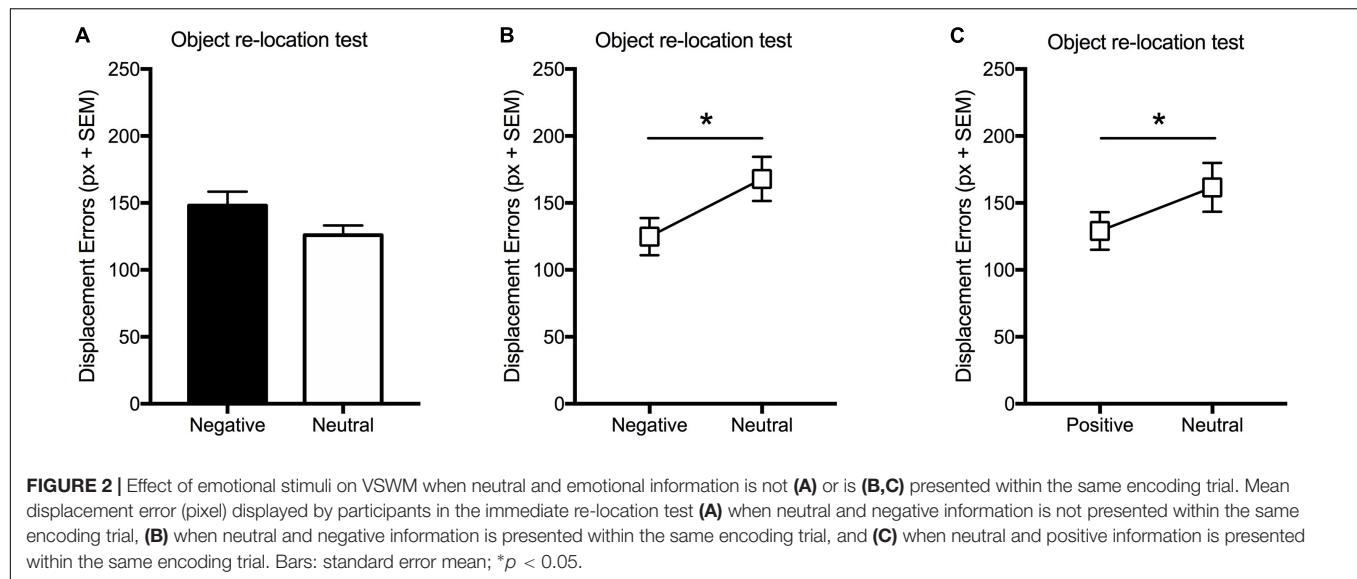
Student's *t*-test and two-way ANOVAs were performed when appropriate on displacement error and memory recall as well as on the level of arousal and valence of pictures. All statistical analyses were performed with SPSS 24 and considering  $\alpha = 0.05$ .

## RESULTS

### Emotional Stimuli Affect Visuospatial Working Memory When They Are in Competition With Neutral Stimuli

The first experiments were carried out in order to verify the competition hypothesis.

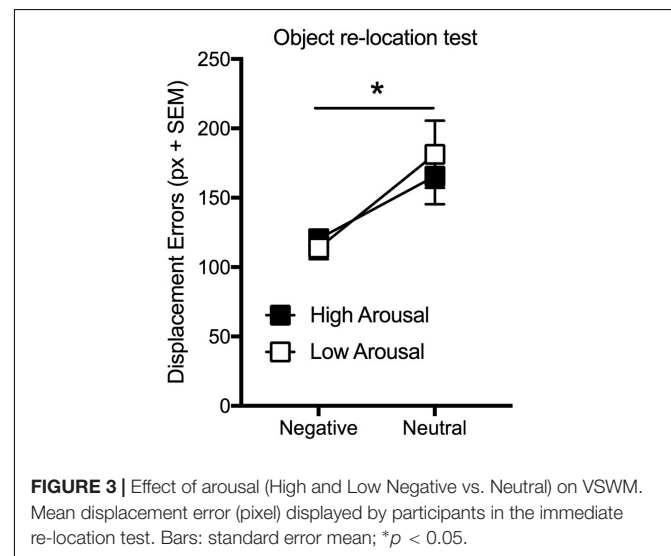
In Experiment 1A object-positions were tagged either by neutral or negative pictures. Thus, the effect of emotion on VSWM was investigated as a between-subject factor. Statistical analysis (unpaired *t*-test) showed no differences [ $t(38) = 1.73$ ;  $p > 0.05$ ] in the relocation performance between groups (**Figure 2A**). Memory for incidentally learned pictures was evaluated in a free recall test carried out 24 h after the object-relocation task. Statistical analysis (unpaired *t*-test) revealed that the number of pictures recalled was significantly [ $t(38) = 2.58$ ;



**TABLE 2 |** Proportion of remembered pictures.

	Mean ( $\pm$ SEM)	
Experiment 1A ( $n = 20 + 20$ )	Negative	Neutral
	0.200 (0.030)	0.096 (0.032)
Experiment 1B ( $n = 28$ )	Negative	Neutral
	0.232 (0.048)	0.116 (0.033)
Experiment 1C ( $n = 25$ )	Positive	Neutral
	0.270 (0.043)	0.140 (0.029)
Experiment 2 ( $n = 29$ )	Negative HIGH arousal	Neutral
Experiment 2 ( $n = 20$ )	0.647 (0.052)	0.405 (0.042)
	Negative LOW arousal	Neutral
	0.425 (0.063)	0.188 (0.051)
Experiment 4 ( $n = 20$ )	Negative	Positive
	0.225 (0.040)	0.238 (0.046)

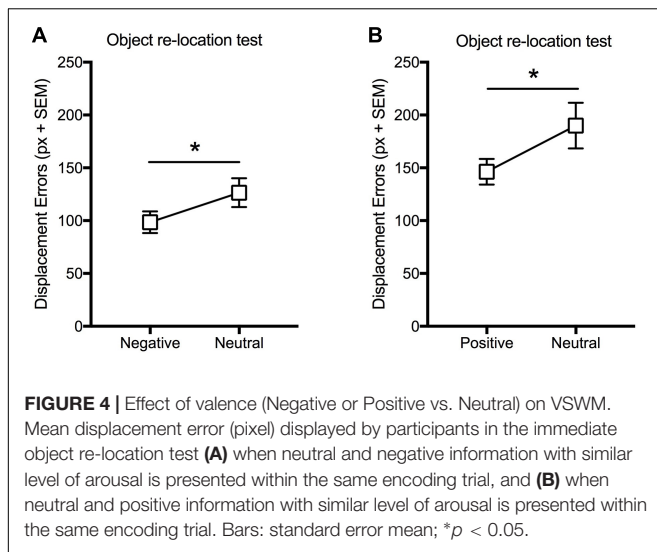
$p < 0.05$ ) greater for negative than for neutral pictures (Table 2). In order to investigate whether a confounding effect of the individual differences in motor control or in spatial working memory ability could influence relocation performance, we carried out a supplemental experiment in which 3 h after the main task performed with negative and neutral IAPS pictures (IAPS pictures condition), participants performed a further object relocation task in which object-positions were tagged by the presentation of pictures built by scrambling pixels of different colors (Control condition). Results were analyzed by means of a two-way ANOVA considering a Group factor with two levels, Negative and Neutral (i.e., participants who were administered negative or neutral IAPS pictures in the first encoding condition), and a Condition factor with two levels, IAPS pictures or Control scrambled pictures. The analysis revealed no significant effects of Group [Negative vs. Neutral;  $F(1,23) = 0.11$ ;  $p = \text{n.s.}$ ], of Condition [IAPS pictures vs. Control pictures;  $F(1,23) = 1.62$ ;  $p = \text{n.s.}$ ] and no interaction between factors [ $F(1,23) = 0.35$ ;  $p = \text{n.s.}$ ; **Supplementary Figure S1**].



In Experiment 1B, object positions during the encoding phase were tagged by the incidental presentation of both neutral and negative pictures, which appeared one at a time. Thus, the effect of emotion on VSWM was investigated as a within-subject factor. Statistical analysis (paired  $t$ -test) carried out on the displacement errors revealed a significant effect [ $t(27) = -2.37$ ;  $p < 0.05$ ] of emotion, with negative-tagged objects being better relocated than neutral-tagged objects (Figure 2B). Like in Experiment 1, the number of remembered pictures in the delayed free recall test was significantly [ $t(27) = 2.22$ ;  $p < 0.05$ ] greater for negative than for neutral pictures (Table 2).

In Experiment 1C, object positions during encoding phase were tagged by the incidental presentation of both neutral and positive pictures, which appeared one at a time. Thus, the effect of emotion on VSWM was investigated as a within-subject factor. Statistical analysis (paired  $t$ -test) carried out on the displacement



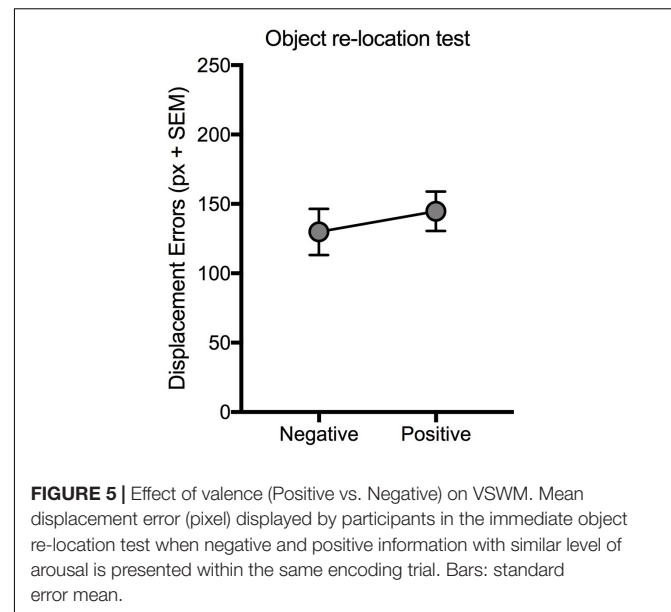


errors revealed objects associated to positive pictures were better relocated than those associated to neutral ones [ $t(24) = -2.45$ ;  $p < 0.05$ ] (Figure 2C). The number of remembered pictures in the delayed free recall test was higher for positive pictures than for neutral ones [ $t(24) = 2.59$ ;  $p < 0.05$ ] (Table 2).

These results confirm the hypothesis that emotional information enhances spatial memory performance when emotional and non-emotional stimuli compete with one another for accessing the working memory system (competition effect). Moreover, the emotional content of images increased long-term memory for the incidentally learned pictures. Since negative and positive pictures were more activating than neutral ones, a possible effect of emotional arousal can be envisaged which allows emotional pictures to get a priority access to the memory system (Mather and Sutherland, 2011). However, since negative and positive pictures are also emotionally valenced, in comparison to neutral ones, a possible role of valence in determining a priority access to the memory system cannot be ruled out. In this case, it is possible to hypothesize a “dichotomic system” in which competition is driven by what is endowed with an emotional valence (either negative or positive) and what is not (neutral). The following experiments are planned to verify the role of “arousal” and “valence” in the emotion-enhancing spatial memory performance in competition condition.

### Increasing Arousal of Negative Pictures Did Not Enhance Spatial Working Memory Performance in the Competition Condition

Experiment 2 was specifically aimed at investigating the role of emotional arousal in the competition mechanism. The ABC theory predicts that the level of arousal of emotional stimuli regulates the access to the working memory system, enhancing the encoding of within-object characteristics, like spatial position (Mather and Sutherland, 2011). Therefore, we expected that increasing the arousal level of negative pictures would lead to



an enhancement of spatial working memory performance. In the high-arousal group, rectangles were tagged by the presentation of neutral and high-arousal negative pictures while in the low-arousal group, rectangles were tagged by neutral and negative pictures with low arousal values (Table 1).

Statistical analysis (two-way ANOVA) carried out on the displacement errors considering arousal (high vs. low) as between factor and valence (negative vs. neutral) as within factor revealed that the effect of valence was significant [ $F(1,47) = 15.71$ ;  $p < 0.01$ ] whereas neither the effect of arousal [ $F(1,47) = 0.06$ ;  $p = 0.62$ ] nor the interaction [ $F(1,47) = 0.62$ ;  $p = 0.43$ ] were significant (Figure 3). Interestingly, in the free recall test, the number of both neutral and negative pictures correctly recalled was significantly greater for the high-arousal group than for the low-arousal one [Two-way ANOVA: arousal effect:  $F(1,47) = 13.88$ ,  $p < 0.005$ ; valence effect:  $F(1,47) = 28.38$ ,  $p < 0.005$ ; interaction effect:  $F(1,47) = 0.002$ ,  $p = 0.96$ ] (Table 2).

These results indicate that arousal manipulation, obtained by selecting negative pictures with different level of arousal (high vs. low), did not lead to a significant enhancement in spatial working memory performance, whereas it did enhance long-term memory for both negative and neutral pictures.

Since arousal did not seem to significantly impact on the competition between negative and neutral information for accessing the working memory system, we explored the possible effect of valence in this competition.

### Valence Affects Spatial Working Memory When Arousal Is Kept Constant in Competition Condition

In Experiment 3, we sought to better ascertain the role of valence in spatial working memory by keeping the level of pictures' arousal constant, and by modulating the level of pictures' valence. Thus, in Experiment 3A we tagged the object

position with negative and neutral pictures but this time pictures had comparable arousal levels (Table 1). In Experiment 3B, rectangle position was tagged by positive and neutral pictures with comparable arousal values (Table 1). In both cases valence is a within-subject factor.

In Experiment 3A, the statistical analysis (paired *t*-test) carried out on displacement errors (Figure 4A) revealed a significant effect of valence [ $t(17) = -2.89$ ;  $p < 0.05$ ] with lower displacement errors for negative-related objects.

In Experiment 3B, the statistical analysis (paired *t*-test) carried out on displacement errors (Figure 4B) revealed that objects associated to positive pictures were better relocated than those associated to neutral ones [ $t(20) = 2.28$ ;  $p < 0.05$ ] (Figure 4B).

The results of these experiments indicate that when arousal is kept constant between neutral and negative pictures, or between positive and neutral pictures, valence significantly affects visuo-spatial performance.

## The Effect of Emotion on Working Memory Performance Vanishes When all Stimuli Have an Emotional Value

Experiment 4 was designed in order to verify the competition hypothesis when neutral stimuli are not presented, and competition within the same trial is only among emotional pictures with different valence (positive and negative) values. Thus, object positions were tagged by negative and positive pictures, with comparable arousal levels (Table 1). Statistical analyses (paired *t*) carried out on displacement errors (Figure 5) and on long-term memory (Table 2) revealed no significant effects of valence [ $t(19) = -0.7$ ;  $p > 0.05$  and  $t(19) = 0.24$ ;  $p > 0.05$ , respectively]. Interestingly, the displacement error of both positive- and negative-related objects ( $144.74 \pm 14.18$  and  $129.84 \pm 16.65$ ) in this experiment was similar to displacement error of neutral-related objects ( $125.95 \pm 9.04$ ) in Experiment 1A, i.e., in absence of competition [one-way ANOVA;  $F(2,57) = 0.56$ ;  $p = 0.58$ ], suggesting that when positive and negative stimuli compete for accessing the working memory the “competition effect” disappears.

In the delayed free recall test, both positive and negative pictures of this experiment were better remembered ( $0.23 \pm 0.04$  and  $0.24 \pm 0.05$ ) than neutral pictures ( $0.12 \pm 0.03$ ) of Experiment 1B [one-way ANOVA;  $F(2,65) = 3.2$ ;  $p < 0.05$ ]. Notably, in this case, both positive and negative pictures had a higher arousal compared to the neutral pictures of Experiment 1B.

Taken together the results of this experiment indicate that when competition occurs among stimuli that all have an emotional value (either positive or negative), the effect of emotion on working memory performance vanishes.

## DISCUSSION

The present study was aimed at investigating the effect of the incidental presentation of emotional stimuli on VSWM. To this purpose an object-relocation task was used in which emotional stimuli appeared superimposed on the objects to be relocated.

The first hypothesis tested was that the emotional content of stimuli affects memory for object position only when neutral and emotional stimuli are in competition with one another, i.e., when they are presented within the same encoding trial. Previous studies on the effect of emotion on spatial working memory have yielded contrasting results: positive, negative or no effects of processing stimuli with an emotional content on spatial working memory performance have been reported (Putman et al., 2004; Dolcos and McCarthy, 2006; Mather et al., 2006; Mather and Nesmith, 2008; Schmidt et al., 2011; Bannerman et al., 2012; Gonzalez-Garrido et al., 2015). For example, there is evidence showing that emotions have a negative impact on spatial working memory because the emotional content of stimuli captures attention, subtracting resources for processing other within-object features, like spatial location (Mather et al., 2006). The possibility that the attentional bias exerted by emotional stimuli distracts from the main task, thus impairing memory performance, has been suggested (Dolcos and McCarthy, 2006; Bannerman et al., 2010; Iordan et al., 2013; Tavares et al., 2016). Although emotional stimuli are able to capture attention, there are studies showing no effect or even an improving effect of emotion on spatial working memory performance (Bannerman et al., 2012; Gonzalez-Garrido et al., 2015). Bannerman et al. (2012) found no effect of emotions on spatial working memory evaluated with a Corsi-Block Task in which spatial positions were highlighted by the presentation of emotional pictures. The authors explained the lack of effect by considering the possibility that the attentional bias exerted by the emotional content of stimuli is specifically directed to object-identity recognition (i.e., what the stimulus is). Therefore, since “what” the stimulus is and “where” the stimulus is located are encoded by different processes, spatial memory could not be affected by the emotional content of stimuli (Bannerman et al., 2012). Gonzalez-Garrido et al. (2015) recorded EEG during a face relocation task and found that spatial memory for happy face position, as well as the electrophysiological component linked to the attentional process, is increased for emotional faces. The authors suggested that emotional faces attracted more attention increasing spatial memory performance through a domain-general attention-based mechanism (Gonzalez-Garrido et al., 2015).

In the present study, the results of the first set of experiments seem to indicate that the emotional content of pictures enhanced VSWM only when emotional and neutral information was processed within the same trial, suggesting the existence of a “competition effect.” Although the interpretation of our results must be modulated by the fact that Experiment 1A entailed a between group comparison with a low level of power, it appears that, if both neutral and negative stimuli have been encountered in the environment (like in Experiment 1B), then the position of the latter is better remembered. Instead, if stimuli encountered all had a similar emotional impact (like in our Experiment 1A), the enhancing effect on spatial working memory performance, observed in the competition condition, disappears. The enhancing effect of emotion on spatial working memory performance in competition condition has been replicated also with neutral and positive stimuli (Experiment 1C).

These results seem to be in line with the ABC theory (Mather and Sutherland, 2011). According to the latter, the arousal level of emotional stimuli should increase the strength of mental representations for emotional material at the expense of the non-emotional one, through an “ABC” process. This “ABC” begins with perception, increasing the perceptual capability for binding features (such as object location) when stimulus competition occurs. The advantage yielded by arousing stimuli is maintained in the working memory system, where emotionally tagged information dominates the competition for mental resources (Richter-Levin and Akirav, 2003; Mather and Sutherland, 2011; Lee et al., 2014; Dunsmoor et al., 2015; Hur et al., 2017; Schweizer et al., 2019). The advantage exerted by emotion in “competition” condition could explain the improving effect of the emotional content of stimuli on spatial working memory performance observed in Experiments 1B and 1C.

In the absence of competition, when all stimuli have a similar emotionally arousing content, like in Experiment 1A, it is plausible to hypothesize that the emotional content of stimuli captures the attentional resources, facilitating the encoding of within-object features. However, when competition between the mental representation of emotional and non-emotional stimuli is lacking (due to the fact that all the stimuli have a similar emotional impact), no prioritization effect emerges during stimulus processing in working memory. This could explain why emotion did not affect spatial working memory performance in Experiment 1A.

To the best of our knowledge, this is the first time that the competition hypothesis is evaluated by comparing the results of two experimental manipulations in which the emotional stimuli compete or not for accessing to working memory. The results of our experiments indicate that competition among stimuli is an important factor to consider when the effect of emotion on memory performance is investigated. Having ascertained that competition is necessary in order for emotions to influence VSWM performance, the question arises as to which dimension, arousal and/or valence, is mainly involved in modulating working memory performance.

In Experiment 2 we sought to verify the prediction that arousal is the emotional dimension that mainly determines the enhancement in spatial working memory performance for the relocation of negative-related objects. A better location memory has been found for arousing pictures independently of their valence, suggesting that arousal, rather than valence, is the critical dimension for the emotion-enhancing memory effect (Mather and Nesmith, 2008). Therefore, we hypothesized that increasing the arousal level of negative pictures should lead to an improvement in spatial memory performance. Overlapping performances in relocating objects associated with high- and low-arousing pictures were observed (Experiment 2), although both high and low arousing negative-related objects were better relocated than neutral-related objects. Even if we cannot definitely rule out the possibility that this pattern of results is linked to a floor effect, our findings do not seem to be in line with the prediction that arousal is the main

emotional dimension affecting VSWM. On the other hand, like for previous experiments, we observed a better memory performance for negative pictures than for neutral ones in the delayed (24 h) free recall task. Moreover, in the second experiment we found that high arousing negative pictures were better remembered than low arousing ones. The latter results support the notion that arousal is mainly involved in long-term memory formation (Bradley et al., 1992; McGaugh, 2000; LaLumiere et al., 2017). Interestingly, the memory improvement effect exerted by the *high arousal condition* was extended also to neutral pictures: neutral pictures presented together with high arousing negative pictures were better remembered than neutral pictures encoded together with low arousing negative pictures.

Two distinct neuronal pathways in the brain have recently been observed for the processing of arousal and valence. Arousing information is processed by an amygdala – hippocampus circuit, while valenced non-arousing information is processed by a PFC-hippocampus circuit, indicating two separate routes for arousal and valence (Kensinger and Corkin, 2004). The former reflects a relatively automatic effect of emotion on long-term memory (especially on memory consolidation), while the latter is associated with a controlled encoding process which supports working memory processes (such as, elaboration or rehearsal of information). In the last experiments we thus sought to further ascertain the role of valence on spatial working memory. This was achieved by presenting emotional pictures with different valence (neutral vs. negative; neutral vs. positive; positive vs. negative) and a similar level of arousal into a single encoding trial.

Findings from Experiment 3 overall show that when arousal is kept constant (at a low level) the position of negative-related objects (Experiment 3A) or positive-related object (Experiment 3B) are better relocated than neutral ones in an immediate VSSP working memory test.

Differently, when both positive and negative stimuli are encoded within the same trial (Experiment 4), there is no effect of emotional valence on spatial performance. In the delayed free recall test, both negative and positive pictures were better remembered than neutral pictures of previous experiments, further supporting the hypothesis that arousal is mainly involved in long-term memory formation. The condition in which both negative and positive pictures are presented but neutral ones are completely absent mimics a condition in which in fact there is no competition since all pictures have a valence. The latter interpretation is supported by the fact that the displacement error of neutral-related objects in Experiment 1A (no competition) is similar to the displacement error of both positive- and negative-related objects in Experiment 4. A possible explanation of our results is that competition for accessing the working memory system could depend upon a dichotomic evaluation between what is endowed with an emotional valence and what is not (emotional vs. non-emotional), independently from the “*valence direction*” (negative or positive). Recently, Willinger et al. (2019) performed an fMRI analysis during an emotional face-matching task and found that the processing of valence directly induces changes in the strength of the bidirectional

coupling within a prefrontal-amygdala circuitry (Willinger et al., 2019). Moreover, Kensinger and Schacter (2006) found that independent brain networks specifically processed either positive or negative stimuli. In particular, the ventro-lateral PFC seemed to respond to negative items, whereas the ventro-medial PFC was more engaged during the processing of positive pictures (Kensinger and Schacter, 2006).

Coming back to our results, it is possible to envisage that valenced stimuli, in comparison with neutral stimuli, increase the coupling strength between PFC and limbic structures, favoring memories for valenced information. In this circuit, networks involving the connection between different subregions of PFC and limbic structures may separately process positive and negative valenced stimuli, preventing competition and favoring memory for both negative and positive information. Future experiments should be specifically planned to investigate this issue.

A number of limitations in the present study have to be considered. First, the task used in the present study relied mainly on memory for spatial location and did not require to assign specific objects to these locations: since all to-be-relocated stimuli were rectangles, participants did not have to assign specific objects to spatial locations. It might be interesting to analyze in further experiments the effect of emotions on object-to-position memory. Moreover, different encoding procedures and experimental material should be considered in future experiments. For example, the effects of incidental vs. intentional encoding of the emotional pictures should be compared in order to verify if the advantage due to the competition effect is independent from the encoding procedure. Indeed, a positive correlation between the release of an arousal-related hormone and long-term memory for emotional pictures has been found with intentional, but not with incidental encoding (Preuß et al., 2009). It is interesting to note, however, that no differences due to encoding procedures were found for spatial working memory (D'Argembeau and Van der Linden, 2004).

## CONCLUSION

In conclusion, the results of the present study suggest that the interplay between arousal and valence is crucial in driving information processing into the memory systems. In this interplay, we hypothesized a differential role of valence and arousal in memory processing. Arousal could act in a relatively automatic way, already from the early stages of perception and perhaps prioritizing the access to the memory system. Moreover, information about the arousal level of learned material can contribute to the long-term memory consolidation process, by the release of arousal-related hormones (i.e., cortisol). In the meanwhile, emotionally valenced stimuli could drive the competition between stimulus representations in the working memory system at a more explicit level. Therefore,

within-object features (such as spatial location) of emotionally valenced stimuli (either positive or negative) could be better processed than those of neutral stimuli, and VSWM performance would be enhanced.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were carried out according to the recommendations of the Committees for Ethics, Department of Psychology Sapienza, University of Rome, and of CERS, LUMSA University. The participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MC developed the idea for this study and drafted the manuscript. MC, DS, CR-A, and VC contributed to the conception and designed the study. BC collected the data and organized the database. MC, CR-A, DS, and VC analyzed and interpreted the data. SL and FD contributed to the discussion of content-related issues and critical revision of the manuscript, and wrote sections of the manuscript. MC and CR-A wrote the final version of the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

## FUNDING

This work was supported by the LUMSA (Fondo di Ateneo per la Ricerca to MC) and Sapienza (Ateneo 2018 to FD).

## ACKNOWLEDGMENTS

We thank Stefano Saraulli for adapting the software for the computer object-relocation task.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02587/full#supplementary-material>

**FIGURE S1** | Mean displacement error (pixel) in a re-location task in which object-positions were tagged by negative (black square) or neutral (white square) IAPS pictures, and in a task (performed 3 hours later) in which object-positions were tagged by pictures built by scrambling pixels of different colors (Control). Bars: standard error mean.



## REFERENCES

- Adelman, J. S., and Estes, Z. (2013). Emotion and memory: a recognition advantage for positive and negative words independent of arousal. *Cognition*. 129, 530–535. doi: 10.1016/j.cognition.2013.08.014
- Anderson, A. K., Wais, P. E., and Gabrieli, J. D. (2006a). Emotion enhances remembrance of neutral events past. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1599–1604. doi: 10.1073/pnas.0506308103
- Anderson, A. K., Yamaguchi, Y., Grabski, W., and Lacka, D. (2006b). Emotional memories are not all created equal: evidence for selective memory enhancement. *Learn. Mem.* 13, 711–718. doi: 10.1101/lm.388906
- Bannerman, L., Milders, M., and Sahraie, A. (2010). Attentional bias to brief threat-related faces revealed by saccadic eye movements. *Emotion* 10, 733–738. doi: 10.1037/a0019354
- Bannerman, L., Temminck, E. V., and Sahraie, A. (2012). Emotional stimuli capture spatial attention but do not modulate spatial memory. *Vision Res.* 65, 12–20. doi: 10.1016/j.visres.2012.05.011
- Bradley, M. M., Greenwald, M. K., Petry, M. C., and Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *J. of Exp. Psychol.* 18, 379–390. doi: 10.1037//0278-7393.18.2.379
- Buchanan, T. W., and Lovall, W. R. (2001). Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology* 26, 307–317. doi: 10.1016/s0306-4530(00)00058-5
- Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., and Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *J. Neurosci.* 20:RC99.
- Cestari, V., Costanzi, M., Castellano, C., and Rossi-Arnaud, C. (2006). A role for ERK2 in reconsolidation of fear memories in mice. *Neurobiol. Learn. Mem.* 86, 133–143. doi: 10.1016/j.nlm.2006.01.003
- D'Argebeau, A., and Van der Linden, M. (2004). Influence of affective meaning on memory for contextual information. *Emotion* 4, 173–188. doi: 10.1037/1528-3542.4.2.173
- Dolcos, F., and Cabeza, R. (2002). Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. *Cogn. Affect. Behav. Neurosci.* 2, 252–263. doi: 10.3758/cabn.2.3.252
- Dolcos, F., LaBar, K. S., and Cabeza, R. (2004). Interaction between the amygdala and the medial temporal lobe memory system predicts better memory for emotional events. *Neuron* 42, 855–863. doi: 10.1016/s0896-6273(04)00289-2
- Dolcos, F., and McCarthy, G. (2006). Brain systems mediating cognitive interference by emotional distraction. *J. Neurosci.* 26, 2072–2079. doi: 10.1523/jneurosci.5042-05.2006
- Dunsmoor, J. E., Murty, V. P., Davachi, L., and Phelps, E. A. (2015). Emotional learning selectively and retroactively strengthens memories for related events. *Nature* 520, 345–348. doi: 10.1038/nature14106
- Gonzalez-Garrido, A. A., Lopez-Franco, A. L., Gomez-Velazquez, F. R., Ramos-Loyo, J., and Sequeira, H. (2015). Emotional content of stimuli improves visuospatial working memory. *Neurosci. Lett.* 585, 43–47. doi: 10.1016/j.neulet.2014.11.014
- Hur, J., Jordan, A. D., Dolcos, F., and Berenbaum, H. (2017). Emotional influences on perception and working memory. *Cogn. Emot.* 31, 1294–1302. doi: 10.1080/02699931.2016.1213703
- Jordan, A. D., Dolcos, S., Denkova, E., and Dolcos, F. (2013). Sex differences in the response to emotional distraction: an event-related fMRI investigation. *Cogn. Affect. Behav. Neurosci.* 13, 116–134. doi: 10.3758/s13415-012-0134-6
- Kang, C., Wang, Z., Surina, A., and Lü, W. (2014). Immediate emotion-enhanced memory dependent on arousal and valence: the role of automatic and controlled processing. *Acta Psychol.* 150, 153–160. doi: 10.1016/j.actpsy.2014.05.008
- Kensinger, E. A. (2009). What factors need to be considered to understand emotional memories? *Emot. Rev.* 1, 120–121. doi: 10.1177/1754073908100436
- Kensinger, E. A., and Corkin, S. (2003). Effect of negative emotional content on working memory and long-term memory. *Emotion* 3, 378–393. doi: 10.1037/1528-3542.3.4.378
- Kensinger, E. A., and Corkin, S. (2004). Two routes to emotional memory: distinct neural processes for valence and arousal. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3310–3315. doi: 10.1073/pnas.0306408101
- Kensinger, E. A., Garoff-Eaton, R. J., and Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *J. Mem. Lang.* 54, 99–112. doi: 10.1016/j.jml.2005.05.005
- Kensinger, E. A., Garoff-Eaton, R. J., and Schacter, D. L. (2007). Effects of emotion on memory specificity in young and older adults. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 62, 208–215.
- Kensinger, E. A., and Schacter, D. L. (2006). Processing emotional pictures and words: effect of valence and arousal. *Cogn. Affect. Behav. Neurosci.* 6, 110–126. doi: 10.3758/cabn.6.2.110
- Kensinger, E. A., and Schacter, D. L. (2007). Remembering the specific visual details of presented objects: neuroimaging evidence for effects of emotion. *Neuropsychologia* 45, 2951–2962. doi: 10.1016/j.neuropsychologia.2007.05.024
- Kessels, R. P., Postma, A., and de Haan, E. H. (1999). Object RELOCATION: a program for setting up, running, and analyzing experiments on memory for object locations. *Behav. Res. Methods Instrum. Comput.* 31, 423–428. doi: 10.3758/bf03200721
- La Bar, K. S. (2007). Beyond fear: emotional memory mechanisms in the human brain. *Curr. Dir. Psychol. Sci.* 16, 173–177. doi: 10.1111/j.1467-8721.2007.00498.x
- LaLumiere, R. T., McGaugh, J. L., and McIntyre, C. K. (2017). Emotional modulation of learning and memory: pharmacological implications. *Pharmacol. Rev.* 69, 236–255. doi: 10.1124/pr.116.013474
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual*. Technical Report A-8. Gainesville, FL: University of Florida.
- Lee, T. H., Sakaki, M., Cheng, R., Velasco, R., and Mather, M. (2014). Emotional arousal amplifies the effects of biased competition in the brain. *Soc. Cogn. Affect. Neurosci.* 9, 2067–2077. doi: 10.1093/scan/nsu015
- Levens, S. M., and Phelps, E. A. (2008). Emotion processing effects on interference resolution in working memory. *Emotion* 8, 267–280. doi: 10.1037/1528-3542.8.2.267
- Leventon, J. S., Camacho, G. L., Ramos Rojas, M. D., and Ruedas, A. (2018). Emotional arousal and memory after deep encoding. *Acta Psychol.* 188, 1–8. doi: 10.1016/j.actpsy.2018.05.006
- Lindstrom, B. R., and Bohlin, G. (2011). Emotion processing facilitates working memory performance. *Cogn. Emot.* 25, 1196–1204. doi: 10.1080/02699931.2010.527703
- Mather, M. (2007). Emotional arousal and memory binding: an object-based framework. *Perspect. Psychol. Sci.* 2, 33–52. doi: 10.1111/j.1745-6916.2007.00028.x
- Mather, M., Mitchell, K. J., Raye, C. L., Novak, D. L., Greene, E. J., and Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *J. Cogn. Neurosci.* 18, 614–625. doi: 10.1162/jocn.2006.18.4.614
- Mather, M., and Nesmith, K. (2008). Arousal-enhanced location memory for pictures. *J. Mem. Lang.* 58, 449–464. doi: 10.1016/j.jml.2007.01.004
- Mather, M., and Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspect. Psychol. Sci.* 6, 114–133. doi: 10.1177/1745691611400234
- McGaugh, J. L. (2000). Memory—a century of consolidation. *Science* 287, 248–251. doi: 10.1126/science.287.5451.248
- McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.* 27, 1–28. doi: 10.1146/annurev.neuro.27.070203.144157
- McGaugh, J. L. (2006). Make mild moments memorable: add a little arousal. *Trends Cogn. Sci.* 10, 345–347. doi: 10.1016/j.tics.2006.06.001
- McReynolds, J. R., and McIntyre, C. K. (2012). Emotional modulation of the synapse. *Rev. Neurosci.* 23, 449–461.
- Moran, T. P. (2016). Anxiety and working memory capacity: a meta-analysis and narrative review. *Psychol. Bull.* 142, 831–864. doi: 10.1037/bul0000051
- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nat. Rev. Neurosci.* 9, 148–158. doi: 10.1038/nrn2317
- Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends Cogn. Sci.* 13, 160–166. doi: 10.1016/j.tics.2009.01.006
- Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Curr. Opin. Neurobiol.* 14, 198–202. doi: 10.1016/j.conb.2004.03.015

- Preuß, D., Schoofs, D., and Wolf, O. T. (2009). Associations between endogenous cortisol levels and emotional memory in young women: influence of encoding instructions. *Stress* 12, 379–387. doi: 10.1080/10253890802524592
- Putman, P., Honk, J. W., Kessels, R. P., Mulder, M., and Koppeschaar, H. P. (2004). Salivary cortisol and short and long-term memory for emotional faces in healthy young women. *Psychoneuroendocrinology* 29, 953–960. doi: 10.1016/j.psyneuen.2003.09.001
- Richter-Levin, G. (2004). The amygdala, the hippocampus, and emotional modulation of memory. *Neuroscientist* 10, 31–39. doi: 10.1177/1073858403259955
- Richter-Levin, G., and Akirav, L. (2003). Emotional tagging memory information in the search for neural mechanism. *Brain Res. Brain Res. Rev.* 43, 247–256. doi: 10.1016/j.brainresrev.2003.08.005
- Schmidt, K., Patnaik, P., and Kensinger, E. A. (2011). Emotion's influence on memory for spatial and temporal context. *Cogn. Emot.* 25, 229–243. doi: 10.1080/02699931.2010.483123
- Schumann, D., Bayer, J., Talmi, D., and Sommer, T. (2018). Dissociation of immediate and delayed effects of emotional arousal on episodic memory. *Neurobiol. Learn. Mem.* 148, 11–19. doi: 10.1016/j.nlm.2017.12.007
- Schweizer, S., Satpute, A. B., Atzil, S., Field, A. P., Hitchcock, C., Black, M., et al. (2019). The impact of affective information on working memory: a pair of meta-analytic reviews of behavioral and neuroimaging evidence. *Psychol. Bull.* 145, 566–609. doi: 10.1037/bul0000193
- Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., and Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion* 6, 40–61. doi: 10.1037/1528-3542.6.1.40
- Tavares, T. P., Logie, K., and Mitchell, D. G. V. (2016). Opposing effects of perceptual versus working memory load on emotional distraction. *Exp. Brain Res.* 234, 2945–2956. doi: 10.1007/s00221-016-4697-4692
- Ventura-Bort, C., Low, A., Wendt, J., Moltò, J., Poy, R., Dolcos, F., et al. (2016). Binding neutral information to emotional contexts: Brain dynamics of long-term recognition memory. *Cogn. Affect. Behav. Neurosci.* 16, 234–247. doi: 10.3758/s13415-015-0385-0
- Walker, M. P. (2010). Sleep, memory and emotion. *Prog. Brain Res.* 185, 49–68. doi: 10.1016/b978-0-444-53702-7.00004-x
- Willinger, D., Karipidis, I. I., Beltrani, S., Pietro, S. V. D., Sladky, R., Walitza, S., et al. (2019). Valence-dependent coupling of prefrontal-amygdala effective connectivity during facial affect processing. *eNeuro* 6:ENEURO.0079-19.2019. doi: 10.1523/ENEURO.0079-19.2019
- Wilson, K. M., Russell, P. N., de Joux, N. R., Blakely, M. J., and Helton, W. S. (2016). Negative stimuli may improve visuospatial working memory. *Proc. Hum. Factors Ergon. Soc.* 60, 1419–1423. doi: 10.1016/j.bandc.2011.03.008
- Yoon, K. L., Kutz, A. M., LeMoult, J., and Joormann, J. (2016). Working memory in social anxiety disorder: better manipulation of emotional versus neutral material in working memory. *Cogn. Emot.* 31, 1733–1740. doi: 10.1080/02699931.2016.1257482

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Costanzi, Cianfanelli, Saraulli, Lasaponara, Doricchi, Cestari and Rossi-Arnaud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

### Experiment 1

#### Experiment 1A

Negative IAPS pictures: # 1220, 6244, 9427, 9630, 6555, 9042, 9495, 9611

Neutral IAPS pictures: # 7035, 7060, 7110, 7491, 7490, 2480, 7700, 7234

#### Experiment 1B

Negative IAPS pictures: # 1220, 6244, 9427, 9630

Neutral IAPS pictures: # 7035, 7060, 7110, 7491

#### Experiment 1C

Positive IAPS pictures: # 1710, 7270, 8190, 8500

Neutral IAPS pictures: # 7035, 7060, 7110, 7491

### Experiment 2

Negative IAPS pictures High arousal: #1050, 9940, 6550, 6230

Negative IAPS pictures low arousal: #1220, 6244, 9427, 9630

Neutral IAPS pictures: #7175, 7010, 7006, 7950 and #7035, 7060, 7110, 7491

### Experiment 3

#### Experiment 3A

Negative IAPS pictures: #2722, 2590, 7078, 9220

Neutral IAPS pictures: #2214, 2484, 7011, 7290

#### Experiment 3B

Positive IAPS pictures: #1441, 2091, 2540, 5780

Neutral IAPS pictures: #2214, 2484, 7011, 7290

### Experiment 4

Negative IAPS pictures: #1220, 6244, 9427, 9630

Positive IAPS pictures: #1710, 5480, 5833, 8470



# Composite Face Effect Predicts Configural Encoding in Visual Short-Term Memory

Lilian Azer and Weiwei Zhang\*

Department of Psychology, University of California, Riverside, Riverside, CA, United States

## OPEN ACCESS

### Edited by:

Tatiana Aloï Emmanouil,  
Baruch College (CUNY),  
United States

### Reviewed by:

Qi-Yang Nie,  
Sun Yat-sen University, China  
Timothy Michael Ellmore,  
City College of New York (CUNY),  
United States

### \*Correspondence:

Weiwei Zhang  
weiwei.zhang@ucr.edu;  
wwzhang@ucr.edu

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 29 July 2019

**Accepted:** 22 November 2019

**Published:** 11 December 2019

### Citation:

Azer L and Zhang W (2019)  
Composite Face Effect Predicts  
Configural Encoding in Visual  
Short-Term Memory.  
Front. Psychol. 10:2753.  
doi: 10.3389/fpsyg.2019.02753

In natural vision, visual scenes consist of individual items (e.g., trees) and global properties of items as a whole (e.g., forest). These different levels of representations can all contribute to perception, natural scene understanding, sensory memory, working memory, and long-term memory. Despite these various hierarchical representations across perception and cognition, the nature of the global representations has received considerably less attention in empirical research on working memory than item representations. The present study aimed to understand the perceptual root of the configural information retained in Visual Short-term Memory (VSTM). Specifically, we assessed whether configural VSTM was related to holistic face processing across participants using an individual differences approach. Configural versus item encoding in VSTM was assessed using Xie and Zhang's (2017) dual-trace Signal Detection Theory model in a change detection task for orientation. Configural face processing was assessed using Le Grand composite face effect (CFE). In addition, overall face recognition was assessed using Glasgow Face Matching Test (GFMT). Across participants, holistic face encoding, but not face recognition accuracy, predicted configural information, but not item information, retained in VSTM. Together, these findings suggest that configural encoding in VSTM may have a perceptual root.

**Keywords:** visual short-term memory, Gestalt, holistic face processing, receiver operating characteristic, individual difference

In natural vision, visual scenes often consist of individual items (e.g., trees) and global emergent properties of items as a whole (e.g., forest). These different levels of representations can all contribute to perception (Navon and Norman, 1983; Kimchi, 1992), natural scene understanding (e.g., Greene and Oliva, 2009), sensory memory (Cappiello and Zhang, 2016), visual short-term memory (Brady et al., 2011; Tanaka et al., 2012; Orhan et al., 2013; Nie et al., 2017), and long-term memory (Hunt and Einstein, 1981; Yonelinas, 2002). In addition, there could be significant interactions between these hierarchical representations, for example, enhanced item processing by the global context (Fine and Minnery, 2009; Santangelo and Macaluso, 2013). Despite these various hierarchical representations across perception and cognition, global representations receive considerably less attention in memory research than item representation (e.g., Brady et al., 2011). The present study has thus assessed whether configural information, one kind of global representations, retained in VSTM is related to overall holistic processing in vision.

The representations of global information in VSTM have gained some support in recent years (e.g., Brady and Alvarez, 2015; Nie et al., 2017). For example, experimental manipulation of configural information at retrieval could either impair or facilitate VSTM for item information



(Jiang et al., 2000, 2004; Treisman and Zhang, 2006). Specifically, changes in configural context (e.g., by changing features of non-probed items) at test can impair VSTM for spatial locations (Jiang et al., 2004) and non-spatial features (Vidal et al., 2005; Jaswal and Logie, 2011). In addition, manipulation of configural encoding upon formation of VSTM representations can also affect the later access to stored VSTM contents (Delvenne et al., 2002; Xu, 2006; Gao and Bentin, 2011; Gao et al., 2013; Peterson and Berryhill, 2013). For example, surrounding circles on orientation bars can considerably reduce VSTM storage capacity for orientation information (Delvenne et al., 2002; Alvarez and Cavanagh, 2008), presumably because the surrounding circles severely interrupted configural encoding (Xie and Zhang, 2017). It is highly likely that these effects of perceptual organization on VSTM are a natural extension of configural encoding in perceptual processing. Consistent with this hypothesis, Gestalt cues such as connectedness (Woodman et al., 2003; Xu, 2006), similarity (Peterson and Berryhill, 2013), and closure (Gao et al., 2016) could facilitate grouping of individual items during VSTM encoding, leading to increased storage capacity. In addition, the configural superiority effect (CSE) has demonstrated that individuals' ability to detect a target among distractors is significantly faster in the presence of contextual cues and closure (Nie et al., 2016). In other words, closure of stimuli allows individuals to form Gestalts resulting in rapid detection of the target stimulus and successful inhibition of distractor stimulus. Nonetheless, it is unclear whether holistic encoding, for example configural and holistic encoding as opposed to first-order processing of isolated feature in object and face recognition (Kimchi, 1994; Maurer et al., 2002; Piepers and Robbins, 2012), is related to VSTM for configural information. The present study has thus assessed whether holistic face processing (e.g., Tanaka and Farah, 1993) can predict configural VSTM across participants, using an individual differences approach.

Configural encoding in VSTM was estimated using a change detection task for orientation and a recently developed Receiver Operating Characteristic (ROC) model (Xie and Zhang, 2017). In this task **Figure 1A**, participants memorized five briefly presented orientation bars in a memory array. Following a 1,000-ms delay interval, participants reported whether an orientation bar in a test array contained a new orientation ("new" response) or the old orientation ("old" response) on a 6-point confidence scale, as compared to the corresponding bar presented at the same location in the memory array. These responses were used to construct ROC curves, the function relating the probability of "old" responses on *old* trials (hit rate) to the probability of "old" responses on *new* trials (false alarm rate) using Signal Detection Theory (Macmillan and Creelman, 2005). The resulting ROCs were then fitted with a Dual-Trace Signal Detection ROC model (DTSD, Xie and Zhang, 2017) to quantitatively assess contributions of item-based encoding (i.e., individual orientation) and configural encoding (i.e., the overall shape of all orientation bars) in VSTM (for details of the model, see Xie and Zhang, 2017).

Holistic encoding in face processing was estimated using the composite face effect (CFE) of the Le Grand face task (Le Grand et al., 2001, 2004; Mondloch et al., 2002). Unlike the face inversion task that taps on second-order relational

information in face processing, the CFE is a more robust measure of holistic face processing (Maurer et al., 2002). In this task, two brief displays of composite faces were presented sequentially (**Figure 1B**). Each face consisted of a top half and a bottom half that form a complete face when combined together. Participants reported whether the top halves of the two composite faces were the same (e.g., faces in the right column in **Figure 1C**) or different (e.g., faces in the left column in **Figure 1C**) across the two displays while ignoring the two bottom halves which were always the same across the two displays. Although the two bottom halves were completely task irrelevant, they can empirically and phenomenologically interfere with the same/different judgments of the two top halves (e.g., Le Grand et al., 2004), due to holistic face processing that encodes the top and bottom face halves as an integrated face instead of separate face segments. Orthogonal to the manipulation of the same versus different top face halves, the top and bottom face halves were misaligned (e.g., faces in the top row in **Figure 1C**) on half of the trials and properly aligned (e.g., faces in the bottom row in **Figure 1C**) on the remaining trials. The interference from the irrelevant bottom face halves tends to be reduced for the misaligned condition, as compared to the aligned condition, because the top and bottom halves from the misaligned condition are less likely to be perceived as an integrated face (Le Grand et al., 2001, 2004). This difference in performance between the aligned and misaligned conditions, the CFE, is thus an operational definition of the interference caused by task-irrelevant bottom face halves.

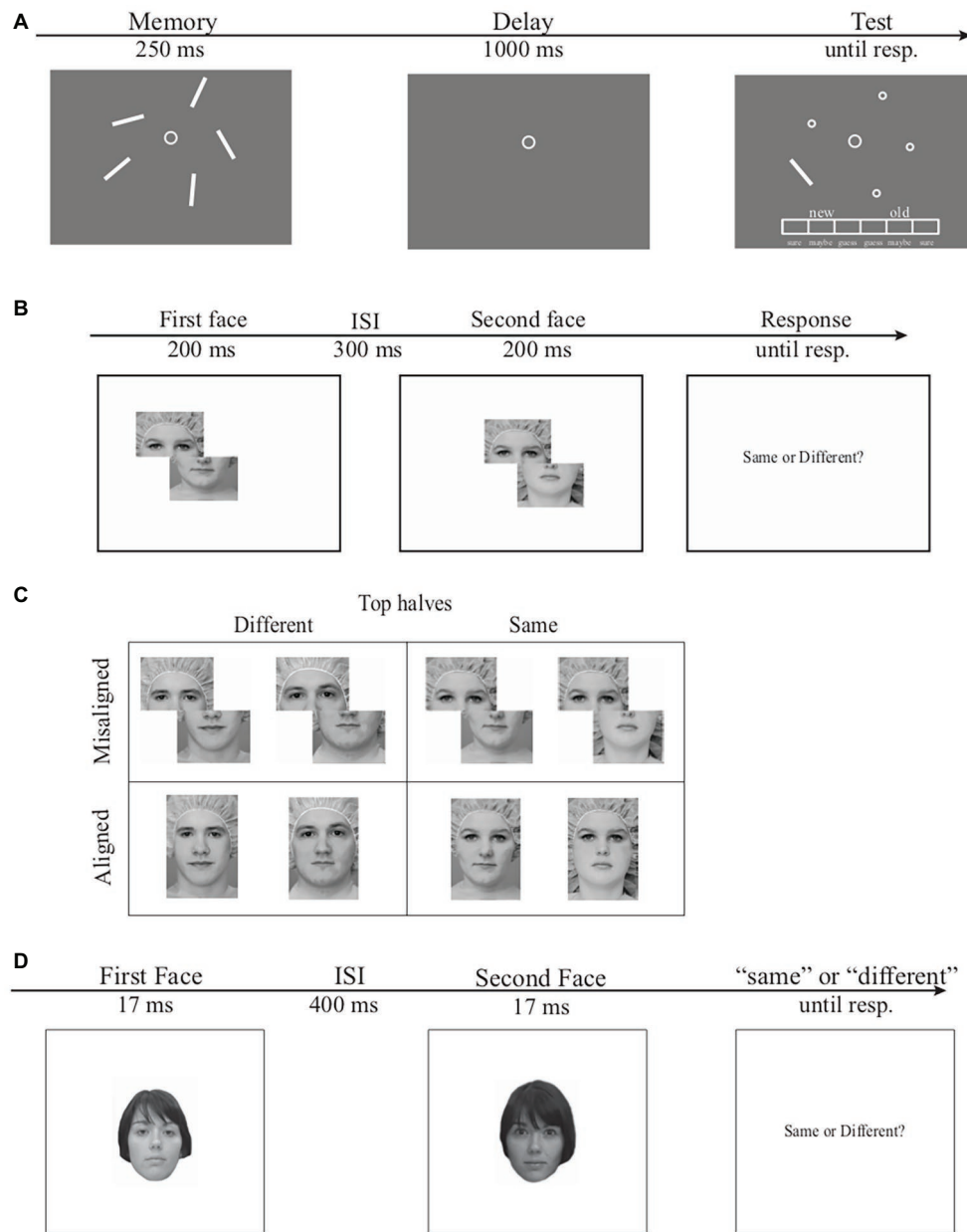
Overall face discrimination was also assessed, using a two-interval forced choice task (**Figure 1D**) with face stimuli from the Glasgow Face Matching Test (GFMT, for details see Burton et al., 2010). Participants in this task reported whether the two sequentially presented faces had the same or different identity. This modified GFMT was chosen over other face identification tasks (Bruce et al., 1999) to minimize potential involvement of VSTM (Xie and Zhang, 2017). Specifically, in the Bruce et al. face identification task for example, a target face is matched to one of the 10 simultaneously presented faces. The matching process in this task could involve several eye movements across stimuli, leading to significant involvements of VSTM (e.g., Irwin, 1991).

We hypothesized that holistic encoding in face processing assessed as CFE, but not the overall face matching ability assessed as accuracy in the modified GFMT discrimination task, would predict configural encoding, but not item encoding, in VSTM across participants.

## METHODS

### Participants

Forty-six UC Riverside students (31 females) between the ages of 18 and 30 with normal color vision and normal or corrected-to-normal visual acuity participated in this study for course credit. Four additional participants were excluded because they did not complete all three tasks within a 1-h experimental



**FIGURE 1 |** Examples of the stimuli and general procedure. **(A)** The procedure and stimulus for the orientation change detection task. On each trial, a memory array of five orientation bars was presented for 250-ms after a 1,000-ms fixation screen. Following the memory display, a 1,000-ms blank delay interval appeared on screen and a test display was presented until participants reported whether the orientation of the probed bar was the same or different from the orientation of the corresponding memory item on a 6-point confidence scale. **(B)** The Le Grand composite-face task. The first face, which was aligned or misaligned, was presented for 200-ms and followed by a 300-ms interstimulus interval. A second face was presented for 200-ms corresponding to the alignment of the first face. Participants were instructed to respond if the top half of the second face was the same or different as the top half of the first face while ignoring the bottom halves. **(C)** Le Grand composite face stimuli. The top row consists of two face pairs from the misaligned condition and the bottom row consists of two face pairs from the aligned condition. The top halves of the two face pairs are either identical to one another (right panel) or different from one another (left panel). **(D)** The modified GFMT task. The first face was presented for 17-ms and followed by a 400-ms interstimulus interval. A second face was presented for 17-ms and participants were instructed to respond if the second face was the same or different as the first face regardless of difference in visual angle or contrast.

session. The experimental procedure was approved by the Institutional Review Board of University of California, Riverside. All participants were provided written informed consent. *A priori* power analysis (Faul et al., 2009) for *r*-based effect

size at a medium level (0.35) suggested that a total sample size of 50 participants would provide 80% statistic power. *Post hoc* power analysis for 46 subjects for a *r*-based effect size of 0.38 yielded 84% statistical power.

## Stimuli and Procedure

All stimuli were presented, using PsychToolbox-3 (Brainard, 1997) for Matlab (The MathWorks, Cambridge, MA), on a LCD monitor with a homogeneous gray background ( $6.7 \text{ cd/m}^2$ ) on a macOS operating system with a refresh rate of 60 Hz at a viewing distance of 57 cm.

The stimuli and procedure were the same as the uncircled bar condition in Experiment 1 of Xie and Zhang (2017). In this VSTM change detection task (**Figure 1A**), the memory array consisted of five white orientation bars ( $3^\circ$  in length and  $0.15^\circ$  in width) in different orientations quasi-randomly selected from  $180^\circ$  circular space. The angular differences between any two orientations were more than  $12^\circ$ . The orientation bars were presented at five locations randomly chosen from eight equally spaced locations on an imaginary circle  $4.5^\circ$  in radius. Each trial began with a 1,000-ms fixation at the center of the screen, followed by a 250-ms memory array of five orientation bars. Participants were required to memorize and retain as many orientation bars as possible over a 1,000-ms blank delay interval. At test, one bar randomly selected from the memory set reappeared at its original location, whereas other memory items were replaced with circles ( $0.3^\circ$ ) as placeholders. Participants reported whether this bar had the “old” or a “new” orientation as compared to the corresponding item at the same location of memory array. The “old”/“new” decision and the confidence for this decision (e.g., sure new, maybe new, or guess new, sure old, maybe old, or guess old) were reported on a 6-point confidence scale ( $16.2$  by  $0.8^\circ$  in visual angle) presented at the bottom of the screen using a computer mouse by the participants. The test orientations were equally likely to be the same as and different from the corresponding memory items. On “new” trials, the orientation was always perpendicular to the original orientation of the memory item ( $90^\circ$  apart). Note, this manipulation rendered mnemonic precision of retained VSTM representations largely irrelevant. That is, either coarse-grain or fine-grained VSTM representation for the test orientation is sufficient for detecting the change between memory and test (for extended discussion, see Xie and Zhang, 2017). Each participant completed 120 trials that were split into three experimental blocks. Responses from this task were fit with the DTSD model, yielding separate estimates of item and configural VSTM encoding for each participant. The details of the DTSD model (e.g., the equations and the theoretical interpretation of the model parameters) and the model fitting procedure were provided in Xie and Zhang (2017), and are thus omitted here.

In the Le Grand face task (**Figure 1B**), each trial began with an 800-ms fixation, followed by sequential presentations of two composite faces with a 300-ms interstimulus interval. Each face was presented for 200-ms. Participants reported whether the top halves of the two sequentially presented faces were the same or different (same and different trials were equally likely), while ignoring the bottom halves, which were always different. Note, to fit the entire experiment within a 1-h session, a partial design was used here, as compared to the complete design that also includes the condition in which the bottom halves were the same (e.g., Richler et al., 2011).

On half of the trials, either the top or bottom halves of the two faces were shifted horizontally to the left by half a face width (the misaligned condition, the top row in **Figure 1C**), whereas on the remaining half of the trials, the top and bottom halves of each face were properly aligned (the aligned condition, the bottom row in **Figure 1C**). The faces in the aligned condition and misaligned condition were presented within a  $4.8^\circ \times 7.2^\circ$  and  $7.2^\circ \times 7.2^\circ$  rectangular area, respectively. The same and different trials were randomly mixed within experimental blocks, whereas misaligned and aligned conditions were blocked with the order counterbalanced across participants. Participants were instructed to make a *Same-or-Different* decision specifically to the two top halves, while ignoring the bottom halves, by pressing button “s” for *Same* or button “d” for *Different* on a computer keyboard as quickly and accurately as possible once the second face appeared. Both response time (RT) and accuracy were recorded. Twenty-five *same* trials and 25 *different* trials were presented for each of the aligned and misaligned conditions, yielding 100 trials in total.

CFE of reaction time (RT) was calculated by subtracting the median RT for misaligned trials from the median RT for aligned trials (Le Grand et al., 2004; Konar et al., 2010). Note, only RTs from trials with correct responses were used in this analysis. CFE was also calculated on mean  $d'$ , a signal detection theory measure (Macmillan and Creelman, 2005), in the same way as CFE on RT (Konar et al., 2010; Richler et al., 2011).

The modified GFMT task is a two-interval forced choice task using face stimuli adopted from Burton and colleagues (Burton et al., 2010). In this task, 150 pairs of gray-scale front-view Caucasian faces, subtending  $5^\circ \times 7^\circ$  of visual angle, were randomly selected from the GFMT set (see Burton et al., 2010 for details). Half of these pairs had matching identities and the other half had different identities. On each trial, two brief displays of faces (17-ms each) were presented sequentially with a 400-ms interstimuli interval. Note, this was different from the original GFMT in which the two faces were presented side by side simultaneously. The sequential presentation in the present study was to match the sequential presentations of face stimuli in the Le Grand task. Participants judged whether the two faces had the same identity or different identities while ignoring visual features that were irrelevant for identities. For example, the two faces with the matching identity in **Figure 1D** had subtle differences in contrasts, hairstyles, face contours, viewing angles, etc. This variance in identity-independent visual features is to ensure participants encode face identities across different views, which mimics faces recognition in natural vision (Burton et al., 2010).

Grubbs' test (Grubbs, 1969) was conducted to detect potential outliers in all measures, although no outlier was identified in the present data, leading to zero outlier rejection.

## RESULTS AND DISCUSSION

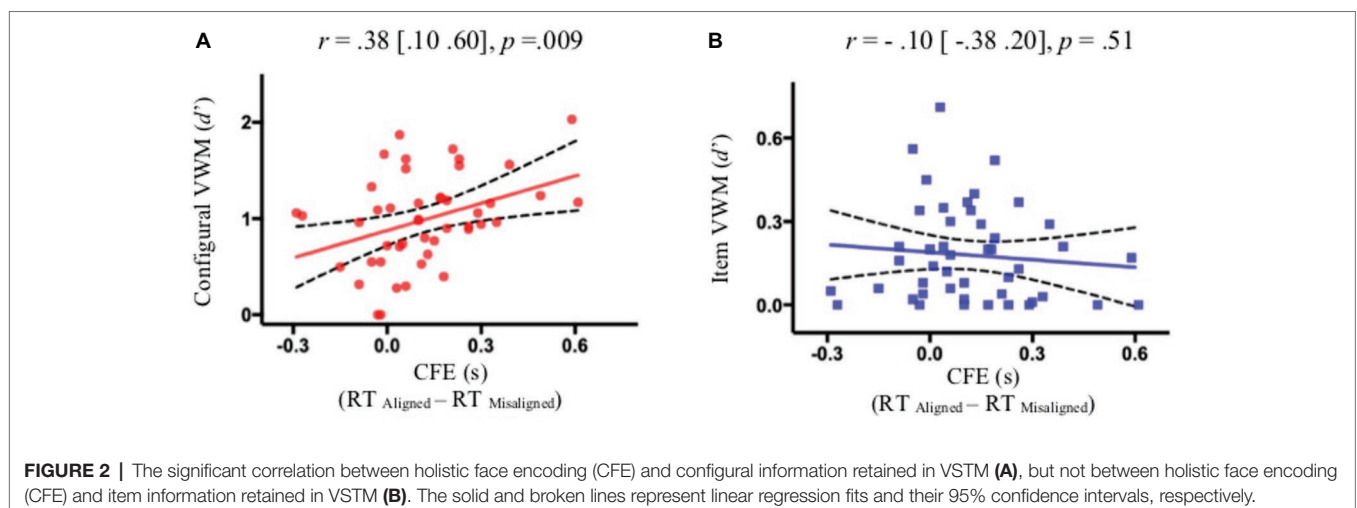
All three tasks were performed with reasonable accuracy. The change detection task yielded an average accuracy of 75% (72%–77%) [mean (95% confidence interval)] and average capacity

(assessed as Cowan's  $K$ , Cowan et al., 2005) of 2.48 (2.23–2.72). For the Le Grand face task, accuracy was averaged at 85% (83%–87%). More importantly, RTs on correct trials were significantly faster on misaligned trials than aligned trials [ $t(45) = 4.23$ ,  $p < 0.001$ , Cohen's  $d = 0.63$ , Bayes factor = 211.96], indicating more interference on aligned trials (and hence significant CFE on RT). Although CFE on  $d'$  was significant [ $t(45) = 4.69$ ,  $p < 0.001$ , Cohen's  $d = 0.70$ , Bayes factor = 818.14], it was not significantly correlated with any measures in the present study ( $p$ 's  $> 0.30$ ) and was thus not discussed further. For the modified GFMT discrimination task, accuracy was averaged at 78% (76%–80%).

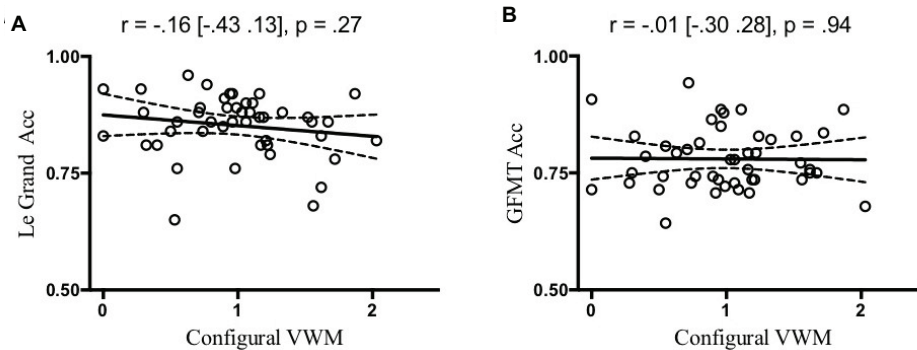
Of central importance, participants with more holistic face processing in the Le Grand face task had larger configural encoding in VSTM (**Figure 2A**), but comparable item encoding (**Figure 2B**), as compared to participants with less holistic face processing. That is, holistic face processing assessed as CFE significantly correlated with configural encoding [Pearson correlation:  $r = 0.38$  (0.10–0.60),  $p = 0.009$ ; Spearman's rank-order correlation for non-Gaussian distribution in RT:  $r = 0.37$  (0.08–0.60),  $p = 0.011$ ], but not with item encoding [Pearson correlation:  $r = -0.10$  (–0.38–0.20),  $p = 0.512$ ; Spearman's rank-order correlation:  $r = -0.10$  (–0.39–0.20),  $p = 0.495$ ]. Additionally, a multivariate regression analysis suggested that CFE significantly predicted configural VWM [ $\beta = 0.95$  (0.25–1.64),  $p = 0.009$ ] but not item VWM [ $\beta = -0.090$  (–0.37–0.19),  $p = 0.512$ ]. Critically, the correlation between CFE and configural encoding was significantly greater than the correlation between CFE and item encoding ( $z = 2.09$ ,  $p = 0.018$ , one-tailed), based on a Fisher's  $r$  to  $z$  transformation one-tailed test of correlated correlation (Meng et al., 1992). The relationship between configural VSTM and holistic face processing seems to be specific in that VSTM configural encoding did not significantly correlate with overall face processing assessed as the accuracy of the Le Grand face task (**Figure 3A**) [Pearson correlation:  $r = -0.16$  (–0.43–0.13),  $p = 0.27$ ; Spearman's rank-order correlation:  $r = -0.25$  (–0.51–0.06),  $p = 0.10$ ] or the accuracy of the modified GFMT (**Figure 3B**) [Pearson correlation:  $r = -0.01$  (–0.30–0.28),  $p = 0.94$ ; Spearman's rank-order correlation:  $r = -0.08$  (–0.37–0.23),  $p = 0.61$ ].

The lack of significant correlation between CFE and the overall face matching ability assessed as the accuracy of 2IFC discrimination task using GFMT face stimuli, although consistent with some previous findings (e.g., Konar et al., 2010), could simply result from the partial design (Richler et al., 2011). As a result, the present study was not optimal for assessing the relationship between configural face encoding and face recognition, which is beyond the scope of the present study.

Interestingly, CFE was significantly correlated with the overall performance of the VSTM change detection task [Accuracy, Pearson correlation:  $r = 0.30$  (0.01–0.54),  $p = 0.04$ ; Spearman's rank-order correlation:  $r = 0.31$  (0.01–0.56),  $p = 0.04$ ; Capacity, Pearson correlation:  $r = 0.30$  (0.01–0.55),  $p = 0.04$ ; Spearman's rank-order correlation:  $r = 0.31$  (0.02–0.56),  $p = 0.03$ ], potentially driven by the significant correlation between the CFE and VSTM configural encoding. The CFE could be considered as the magnitude of distractor processing in a way similar to the flanker compatibility effect (Zhang and Luck, 2014) given that both effects reflect how much distracting information (flanker distractor letters in the flanker task or bottom face halves in the Le Grand composite-face task) is processed. However, the positive correlation between VSTM capacity and the CFE seems to be inconsistent with the load theory of attention which predicts that higher WM capacity (equivalent to low cognitive load) would reduce distractor processing (for a recent review, see de Fockert, 2013; but see, Konstantinou and Lavie, 2013; Zhang and Luck, 2014). In other words, individuals with lower working memory capacities process distractors in an equivalent manner as higher working memory load conditions in tasks investigating the load theory of attention (de Fockert, 2013). In the present study, higher VSTM accuracy and capacity predicted a higher CFE, which is indicative of greater distractibility in the aligned face condition compared to the misaligned face condition. This inconsistency may result from the holistic nature of face processing in the Le Grand task. Further research is needed to directly compare effects of cognitive load on CFE and flanker compatibility effects.







**FIGURE 3 |** The correlation between configural VWM and Le Grand face task accuracy (A) and GFMT task accuracy (B) was not significant. The solid and broken lines represent linear regression fits and their 95% confidence intervals, respectively.

## GENERAL DISCUSSION

Given the hierarchical nature of visual representations in natural vision, it is essential for VSTM to retain hierarchical representations such as item and configural information. The present study assessed configural encoding and item encoding in VSTM for orientation using ROC modeling of change detection performance and assessed holistic face processing using the Le Grand CFE. We found that configural encoding, but not item encoding, for orientations in VSTM significantly correlated with holistic processing in face discrimination in that participants with more configural VSTM also showed larger holistic face encoding. These findings add to the growing literature on the effects of perceptual organization on VSTM.

In addition to its selective correlation with holistic face processing, configural information retained in VSTM could also be experimentally dissociated from item information retained in VSTM using selective experimental manipulations (Xie and Zhang, 2017). Together these findings are essential for scaling up models for VSTM to account for both item information and global stimulus structures (Kroll et al., 1996; Reinitz et al., 1996), especially for natural stimuli (Greene and Oliva, 2009; Brady et al., 2011; Brady and Tenenbaum, 2013).

The Dual Trace Signal Detection model for VSTM is mathematically equivalent to the Dual-Process Signal Detection (DPSD) of Recognition memory (Yonelinas et al., 1999). According to DPSD model, recognition is based on a high-threshold discrete recollection and a continuous familiarity (modeled as  $d'$ ). The  $d'$  component shared across the two models seems to suggest some relationship between configural encoding and familiarity. For instance, configural information may be more important for supporting familiarity than item information. Consequently, face inversion that profoundly impairs holistic face processing (for review, see Maurer et al., 2002) can significantly reduce familiarity (Yonelinas et al., 1999). In addition, processing of higher level representations (e.g., summary statistics and configural information) is more automatic than processing of discrete item representations (e.g., Alvarez and Oliva, 2009), consistent

with the proposal that familiarity is less controlled than recollection (Yonelinas and Jacoby, 1996; Yonelinas, 2002).

In the present study, configural VSTM correlates with holistic face encoding (CFE), but not with overall face discrimination (the 2IFC using GFMT stimuli). The latter finding was not necessarily inconsistent with a previously reported correlation between VSTM and face identification (Megreya and Burton, 2006). The face identification task in Megreya and Burton's (2006) study used a procedure in which the participants match one target face to one of 10 possible faces. This procedure could exert a high demand on VSTM. When demand on VSTM was significantly reduced using a two-alternative force choice task with two simultaneously presented faces in the GFMT, no significant correlation was found between face matching and VSTM (Burton et al., 2010). It is thus possible that a robust relationship between VSTM and face processing could manifest to face recognition tasks with high working memory load.

The present study has focused on how configural processing is shared between face perception and VSTM for orientation. This issue is orthogonal to the role of holistic processing in face processing (Konar et al., 2010; Richler et al., 2011; Gold et al., 2012; DeGutis et al., 2013), which is beyond the scope of the present study. While Konar et al. (2010) used a full factorial design in which the bottom face halves were equally likely to be the same or different the current study used a partial design. The partial design may be suboptimal in that it could induce systematic biases of reporting "same," which may account for some conflicting findings on the effects of holistic encoding in face identification (Konar et al., 2010; Richler et al., 2011). However, it is less of an issue in the present study given a significant relationship was found between holistic face processing and configural VSTM. Nonetheless, it is important for future research to replicate this finding using a full factorial design in the Le Grand CFE task. It will also be interesting to link item and configural encoding in VSTM to part-based encoding and holistic encoding in processing of facial expressions (Tanaka et al., 2012). Furthermore, a stronger test of the relationship between holistic face processing and configural encoding in VSTM is to assess whether holistic face processing can predict increased VSTM performance in VSTM

from a condition where configural encoding is minimized (Xie and Zhang, 2017) to conditions where configural encoding is prominent (Xie and Zhang, 2017).

In summary, the present study, using an individual differences approach, has demonstrated that VSTM for configural information can be accounted for by holistic face perception, providing preliminary evidence that configural encoding in VSTM may result from configural encoding in perception.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of California, Riverside IRB. The

patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

WZ designed the study, collected data, and conducted data analysis. Both authors contributed to manuscript preparation.

## FUNDING

This study was made possible by funding support from the National Institute of Mental Health (R01MH117132 and 1F32MH090719 to WZ).

## ACKNOWLEDGMENTS

We would like to thank Mike Burton for sharing the GFMT face stimuli.

## REFERENCES

- Alvarez, G. A., and Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Percept. Psychophys.* 70, 346–364. doi: 10.3758/PP.70.2.346
- Alvarez, G. A., and Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc. Natl. Acad. Sci. USA* 106, 7345–7350. doi: 10.1073/pnas.0808981106
- Brady, T. F., and Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *J. Vis.* 15, 6–24. doi: 10.1167/15.15.6
- Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vis.* 11, 4–4. doi: 10.1167/11.5.4
- Brady, T., and Tenenbaum, J. (2013). A probabilistic model of visual working memory: incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* 120, 85–109. doi: 10.1037/a0030779
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., and Miller, P. (1999). Verification of face identities from images captured on video. *J. Exp. Psychol. Appl.* 5, 339–360. doi: 10.1037/1076-898X.5.4.339
- Burton, A. M., White, D., and McNeill, A. (2010). The glasgow face matching test. *Behav. Res. Methods* 42, 286–291. doi: 10.3758/BRM.42.1.286
- Cappiello, M., and Zhang, W. (2016). A dual-trace model for visual sensory memory. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1903–1922. doi: 10.1037/xhp0000274
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cogn. Psychol.* 51, 42–100. doi: 10.1016/j.cogpsych.2004.12.001
- de Fockert, J. W. (2013). Beyond perceptual load and dilution: a review of the role of working memory in selective attention. *Front. Psychol.* 4, 1–12. doi: 10.3389/fpsyg.2013.00287
- DeGutis, J., Wilmer, J., Mercado, R. J., and Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition* 126, 87–100. doi: 10.1016/j.cognition.2012.09.004
- Delvenne, J. F., Braithwaite, J. J., Riddoch, M. J., and Humphreys, G. W. (2002). Capacity limits in visual short-term memory for local orientations. *Curr. Psychol. Cogn.* 21, 681–690.
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Fine, M. S., and Minnery, B. S. (2009). Visual salience affects performance in a working memory task. *J. Neurosci.* 29, 8016–8021. doi: 10.1523/JNEUROSCI.5503-08.2009
- Gao, Z., and Bentin, S. (2011). Coarse-to-fine encoding of spatial frequency information into visual short-term memory for faces but impartial decay. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1051–1064. doi: 10.1037/a0023091
- Gao, Z., Ding, X., Yang, T., Liang, J., and Shui, R. (2013). Coarse-to-fine construction for high-resolution representation in visual working memory. *PLoS One* 8:e57913. doi: 10.1371/journal.pone.0057913
- Gao, Z., Gao, Q., Tang, N., Shui, R., and Shen, M. (2016). Organization principles in visual working memory: evidence from sequential stimulus display. *Cognition* 146, 277–288. doi: 10.1016/j.cognition.2015.10.005
- Gold, J. M., Mundy, P. J., and Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychol. Sci.* 23, 427–434. doi: 10.1177/0956797611427407
- Greene, M. R., and Oliva, A. (2009). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.* 58, 137–176. doi: 10.1016/j.cogpsych.2008.06.001
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics* 11, 1–21. doi: 10.1080/00401706.1969.10490657
- Hunt, R. R., and Einstein, G. O. (1981). Relational and item-specific information in memory. *J. Verbal Learn. Verbal Behav.* 20, 497–514. doi: 10.1016/S0022-5371(81)90138-9
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cogn. Psychol.* 23, 420–456. doi: 10.1016/0010-0285(91)90015-G
- Jaswal, S., and Logie, R. (2011). Configural encoding in visual feature binding. *J. Cogn. Psychol.* 23, 586–603. doi: 10.1080/20445911.2011.570256
- Jiang, Y., Chun, M. M., and Olson, I. R. (2004). Perceptual grouping in change detection. *Percept. Psychophys.* 66, 446–453. doi: 10.3758/BF03194892
- Jiang, Y., Olson, I. R., and Chun, M. M. (2000). Organization of visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 683–702. doi: 10.1037/0278-7393.26.3.683
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. *Psychol. Bull.* 112, 24–38. doi: 10.1037/0033-2909.112.1.24
- Kimchi, R. (1994). The role of wholistic/configural properties versus global properties in visual form perception. *Perception* 23, 489–504.
- Konar, Y., Bennett, P. J., and Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychol. Sci.* 21, 38–43. doi: 10.1177/0956797609356508

- Konstantinou, N., and Lavie, N. (2013). Dissociable roles of different types of working memory load in visual detection. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 919–924. doi: 10.1037/a0033037
- Kroll, N. E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., and Tulving, E. (1996). Cohesion failure as a source of memory illusions. *J. Mem. Lang.* 35, 176–196. doi: 10.1006/jmla.1996.0010
- Le Grand, R., Mondloch, C. J., Maurer, D., and Brent, H. P. (2001). Neuroperception. Early visual experience and face processing. *Nature* 410:890. doi: 10.1038/35073749
- Le Grand, R., Mondloch, C. J., Maurer, D., and Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychol. Sci.* 15, 762–768. doi: 10.1111/j.0956-7976.2004.00753.x
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection theory. 2nd Edn.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Maurer, D., Grand, R. L., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn. Sci.* 6, 255–260. doi: 10.1016/S1364-6613(02)01903-4
- Megreya, A. M., and Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Mem. Cogn.* 34, 865–876. doi: 10.3758/BF03193433
- Meng, X.-L., Rosenthal, R., and Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychol. Bull.* 111, 172–175. doi: 10.1037/0033-2909.111.1.172
- Mondloch, C. J., Le Grand, R., and Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception* 31, 553–566. doi: 10.1068/p3339
- Navon, D., and Norman, J. (1983). Does global precedence really depend on visual angle? *J. Exp. Psychol. Hum. Percept. Perform.* 9, 955–965.
- Nie, Q.-Y., Maurer, M., Müller, H. J., and Conci, M. (2016). Inhibition drives configural superiority of illusory Gestalt: combined behavioral and drift-diffusion model evidence. *Cognition* 150, 150–162. doi: 10.1016/j.cognition.2016.02.007
- Nie, Q.-Y., Müller, H. J., and Conci, M. (2017). Hierarchical organization in visual working memory: from global ensemble to individual object structure. *Cognition* 159, 85–96. doi: 10.1016/j.cognition.2016.11.009
- Orhan, A., Jacobs, R., and Jacobs, R. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychol. Rev.* 120, 297–328. doi: 10.1037/a0031541
- Peterson, D. J., and Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychon. Bull. Rev.* 20, 1282–1289. doi: 10.3758/s13423-013-0460-x
- Piepers, D., and Robbins, R. (2012). A review and clarification of the terms “holistic,” “configural,” and “relational” in the face perception literature. *Front. Psychol.* 3, 1–11. doi: 10.3389/fpsyg.2012.00559
- Reinitz, M. T., Verfaellie, M., and Milberg, W. P. (1996). Memory conjunction errors in normal and amnesic subjects. *J. Mem. Lang.* 35, 286–299. doi: 10.1006/jmla.1996.0016
- Richler, J. J., Cheung, O. S., and Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychol. Sci.* 22, 464–471. doi: 10.1177/0956797611401753
- Santangelo, V., and Macaluso, E. (2013). Visual salience improves spatial working memory via enhanced parieto-temporal functional connectivity. *J. Neurosci.* 33, 4110–4117. doi: 10.1523/JNEUROSCI.4138-12.2013
- Tanaka, J. W., and Farah, M. J. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol. A* 46, 225–245.
- Tanaka, J. W., Kaiser, M. D., Butler, S., and Le Grand, R. (2012). Mixed emotions: holistic and analytic perception of facial expressions. *Cognit. Emot.* 26, 961–977. doi: 10.1080/02699931.2011.630933
- Treisman, A., and Zhang, W. (2006). Location and binding in visual working memory. *Mem. Cogn.* 34, 1704–1719. doi: 10.3758/BF03195932
- Vidal, J. R., Gauchou, H. L., Tallon-Baudry, C., and O'Regan, J. K. (2005). Relational information in visual short-term memory: the structural gist. *J. Vis.* 5, 244–256. doi: 10.1167/5.3.8
- Woodman, G. F., Vecera, S. P., and Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychon. Bull. Rev.* 10, 80–87. doi: 10.3758/BF03196470
- Xie, W., and Zhang, W. (2017). Discrete item-based and continuous configural representations in visual short-term memory. *Vis. Cogn.* 25, 21–33. doi: 10.1080/13506285.2017.1339157
- Xu, Y. (2006). Understanding the object benefit in visual short-term memory: the roles of feature proximity and connectedness. *Percept. Psychophys.* 68, 815–828. doi: 10.3758/bf03193704
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517. doi: 10.1006/jmla.2002.2864
- Yonelinas, A. P., and Jacoby, L. (1996). Noncritical recollection: familiarity as automatic, irrelevant recollection. *Conscious. Cogn.* 5, 131–141. doi: 10.1006/ccog.1996.0008
- Yonelinas, A. P., Kroll, N. E., Dobbins, I. G., and Soltani, M. (1999). Recognition memory for faces: when familiarity supports associative recognition judgments. *Psychon. Bull. Rev.* 6, 654–661. doi: 10.3758/BF03212975
- Zhang, W., and Luck, S. J. (2014). Opposite effects of capacity load and resolution load on distractor processing. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 22–27. doi: 10.1037/xhp0000013

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Azer and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Spatial Organization in Self-Initiated Visual Working Memory

Hagit Magen<sup>1\*</sup> and Tatiana Aloï Emmanouil<sup>2</sup>

<sup>1</sup>School of Occupational Therapy, The Hebrew University, Jerusalem, Israel, <sup>2</sup>Program in Psychology, Psychology Department, The Graduate Center, Baruch College, City University of New York, New York, NY, United States

## OPEN ACCESS

### Edited by:

Richard A. Abrams,  
Washington University in St. Louis,  
United States

### Reviewed by:

Jason Rajsic,  
Vanderbilt University,  
United States  
Xuemin Zhang,  
Beijing Normal University,  
China

### \*Correspondence:

Hagit Magen  
msmagen@mail.huji.ac.il

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 August 2019

**Accepted:** 19 November 2019

**Published:** 13 December 2019

### Citation:

Magen H and Emmanouil TA (2019)  
Spatial Organization in Self-Initiated  
Visual Working Memory.  
Front. Psychol. 10:2734.  
doi: 10.3389/fpsyg.2019.02734

Ample research in visual working memory (VWM) has demonstrated that the memorized items are maintained in integrated spatial configurations, even when the spatial context is task irrelevant. These insights were obtained in studies in which participants were provided with the information they memorized. However, the encoding of provided information is only one aspect of memory. In everyday life, individuals often construct their own memory representations, an aspect of memory we have previously termed self-initiated (SI) working memory. In this study, we employed a SI VWM task in which participants selected the visual targets they memorized. The spatial locations of the targets were task irrelevant. Nevertheless, we were interested to see whether participants would construct spatially structured memory representations, which would suggest that they intended to maintain the visual targets as integrated spatial configurations. The results of two experiments demonstrated that participants constructed spatially structured configurations relative to random displays. Specifically, participants selected visual targets in close spatial proximity and constructed spatial sequences with short distances and fewer path crossings. When asked to construct configurations for a hypothetical competitor in a memory contest, participants disrupted the spatial structure by selecting visual targets that were further apart and by increasing the distances between them, which suggests that these characteristics were under their control. At the end of each experiment, participants provided verbal descriptions of the strategies they used to construct the memory displays. While the spatial structure of the SI memory representations was robust, it was absent from the participants' explicit descriptions, which focused on non-spatial strategies. Participants reported selecting items based, most frequently, on semantic categories and visual features. Taken together, these results demonstrated that participants had access to the metacognitive knowledge on the spatial structure of VWM representations, knowledge they manipulated to construct memory representations that enhanced or disrupted memory performance. While having a profound impact on behavior, this metacognitive knowledge on spatial structure remained implicit, as it was absent from the participants' verbal reports. Viewed from a larger perspective, this study explores how individuals interact with the world by actively structuring their surroundings to maximize cognitive performance.

**Keywords:** visual working memory, self-initiation, Gestalt, grouping, metacognition, metamemory



Our everyday interaction with a visually rich and complex world is often aided by short-lived internal representations of relevant information from our surroundings. Visual working memory (VWM) is the mechanism in charge of the formation and temporary maintenance of such representations (see Luck and Vogel, 2013; Ma et al., 2014, for reviews). While rich and complex, the world is also highly structured and governed by spatial regularities such as those captured by the Gestalt organization cues (Wagemans et al., 2012). Consistent with the complexity and inherent structure in our surroundings, the basic representations of VWM are also complex, consisting of interconnected multi-level visual objects (Jiang et al., 2000; Brady et al., 2011; Orhan and Jacobs, 2014). Moreover, memory representations that follow real world regularities typically benefit VWM performance. For instance, visual displays in which items are grouped by any number of Gestalt organization cues such as proximity, similarity, connectedness or symmetry, yield higher accuracy rates relative to unstructured displays (e.g., Woodman et al., 2003; Rossi-Arnaud et al., 2006; Peterson and Berryhill, 2013; Gao et al., 2015; Jiang et al., 2016; van Lamsweerde et al., 2016). Structured displays appear to benefit performance by reducing cognitive and neural loads due to the compression of maintained information into higher order configurations (i.e., chunks, Miller, 1956), that effectively increases memory capacity (Xu and Chun, 2007; Gao et al., 2011; Luria and Vogel, 2014; Peterson et al., 2015).

Spatially structured memory displays, such as those based on Gestalt cues, encourage grouping and consequently the maintenance of independent visual items as integrated spatial configurations. However, other lines of research suggest that the formation and maintenance of spatially integrated memory representations is more fundamental to VWM and occurs even when the encoded visual displays are unstructured and space is overall task irrelevant (Gratton, 1998; Jiang et al., 2000; Treisman and Zhang, 2006). For instance, in a color VWM task in which participants were probed on individual colored items, Jiang et al. (2000) changed the irrelevant locations of the memory targets between encoding and retrieval (i.e., disrupted the overall spatial configuration of the encoded display during retrieval). The disruption of the spatial configuration during the retrieval phase decreased memory performance, suggesting that individual items were encoded and maintained as integrated spatial configurations, even when the displays were spatially unstructured. Thus, space seems to have a unique and fundamental role in VWM.

In the studies reviewed thus far, participants memorized visual displays that were provided to them, and therefore had no control over the memorized content. From these displays, participants extracted and subsequently maintained the overall spatial configuration of the display, a process that is thought to occur quickly and relatively effortlessly (Jiang et al., 2000). The maintenance of provided information, however, is only one aspect of memory performance in everyday life. In many scenarios, memory is self-initiated as individuals shape the content of their own memory representations. For example, individuals often place objects in different locations and retrieve them a short while after. We have recently begun to explore this aspect of

memory, we termed self-initiated (SI) WM, which although is prevalent in everyday behavior, is largely unexplored in the WM literature (Magen and Emmanouil, 2018, 2019; Magen and Berger-Mandelbaum, 2018; Milchgrub and Magen, 2018; Berger-Mandelbaum and Magen, 2019).

An important question regarding SI WM is whether individuals construct memory representations with properties that are consistent with the basic function and structure of memory. Put differently, assuming that individuals select memory representations with an attempt to maximize performance, would they have access to the metacognitive knowledge of the structure of efficient WM representations, knowledge that would allow SI WM to operate efficiently. Given the fundamental role of space in the structure of efficient VWM representations, in the current study we ask whether space is fundamental to SI VWM representations as well. Results from a recent SI VWM study identified a robust spatial structure when space was task relevant and the entire spatial context was constant throughout the experiment (Magen and Berger-Mandelbaum, 2018). The current study takes a further step in understanding the role of space in SI VWM, by testing whether a spatial structure would still be present in SI VWM when space was task irrelevant.

Magen and Berger-Mandelbaum (2018) explored the structure of SI VWM representations, using a modified change detection task. In each trial, participants were presented with a horizontal display of eight visual targets (either real-world objects or abstract shapes) from which they selected three or four targets they memorized and then placed them in several locations in a circular array of eight locations. On half of the trials one of the targets repeated. Following a short delay, participants were probed on object-location conjunctions, deeming space task relevant.

Verbal reports provided by the participants and analysis of the spatial configurations they constructed were used to uncover the strategies that guided them in the construction of the SI VWM representations. The results showed that abstract shapes were selected most frequently based on their resemblance to familiar objects that could be verbalized, while real world objects were mostly selected based on visual features such as color. While participants reported selecting visual targets based on these non-spatial features, their selections were spatially biased to targets presented on the left and central parts of the horizontal target display from which they selected the visual targets they memorized. Importantly, when faced with the circular array, participants placed the to-be memorized visual targets in structured spatial configurations, organized most frequently by the Gestalt organization cue of symmetry and to a lesser extent by cues of proximity and similarity. Participants also formed complex representations, which were based on the interaction of two Gestalt organization cues of proximity and similarity.

Notably, the construction of the SI VWM memory displays was time consuming. Reaction time (RT) for the first visual target or the first location of the sequences participants selected were longer relative to subsequent items in the sequence and increased with set size. These RT findings suggested that

participants invested time in planning the memory displays they constructed before they executed their selections. Overall, the results of Magen and Berger-Mandelbaum (2018) suggested that participants have access to the metacognitive knowledge on the benefit of structure (based on Gestalt cues) in VWM and invested time and resources during encoding to construct spatially structured displays in order to maximize maintenance and retrieval processes.

In the current study, we explored how fundamental is space in the structure of SI VWM representations. Unlike our previous study (Magen and Berger-Mandelbaum, 2018), space was irrelevant during retrieval, and the spatial context varied randomly between trials. We assumed that if participants have a metacognitive knowledge on the spatial structure of representations in VWM, they would invest resources in constructing spatially structured memory representations, although the task emphasized only non-spatial visual information. Note that in this respect, SI VWM deviates considerably from non-SI (provided) VWM. While the spatial configuration is an emergent property that is easily extracted from the visual display when the memory displays are provided to the participants (Jiang et al., 2000), building such representations is time consuming (Magen and Berger-Mandelbaum, 2018; Magen and Emmanouil, 2018). Moreover, the visual targets in the current study were distributed randomly across the display, and therefore imposing a spatial structure on the selected targets could potentially constrain the use of non-spatial strategies.

Our main analysis in this study focused on the spatial and non-spatial characteristics of the memory representations that participants constructed (see section “Experiment 1” for details). Construction of spatially structured memory representations would suggest that participants have access to the metacognitive knowledge on the fundamental role of space in VWM. Nevertheless, the construction of such representations would not reveal whether that knowledge is implicit or explicit. A strategy questionnaire that participants filled out and manipulations introduced in Experiment 2 explored the extent to which this metacognitive knowledge was explicit and could be strategically manipulated.

## EXPERIMENT 1

The goal of Experiment 1 was to examine whether participants would construct spatially structured memory representations in a SI VWM task, in which space was task irrelevant. Participants were presented on each trial with displays of 12 randomly distributed pictures of real world objects. In the SI encoding condition, they were asked to select 1–7 pictures to memorize. An additional non-SI (i.e., provided) condition was introduced in the task, in which participants memorized 1–7 pictures that were randomly selected for them by the computer. Following a short delay, participants were probed on a single central target and indicated whether it matched or not one of the memorized items. The spatial structure formed between the targets selected in the SI condition was evaluated and was compared to the spatially unstructured non-SI representations.

As in our previous studies (Magen and Emmanouil, 2018; Milchgrub and Magen, 2018), spatial structure was defined based on a body of literature on provided (non-SI) spatial WM, which had identified the main characteristics of structured spatial configurations that benefited memory performance. Because these characteristics are relevant for the current study, we describe them here in detail. Note that thus far we have used the term spatial configuration to describe the spatial structure inherent in memory displays. In the context of the present study (and previous studies on spatial WM), we will also use the term spatial sequence to capture the dynamic construction process of the spatial configurations.

The literature has shown that structured spatial sequences that were based on familiar shapes, or followed well-established perceptual Gestalt organization cues of proximity, good continuation, symmetry, and linearity benefited memory performance (Kemps, 2001; Bor et al., 2003; De Lillo, 2004; Parmentier et al., 2005). Specifically, two characteristics of the spatial sequence path, an imaginary line between two successive to-be-remembered locations in the sequence, were shown to have an impact on memory performance. One of these characteristics is the path length, defined as the distance between two successive locations in the sequence. Sequences with longer paths were correlated with poorer memory performance, a finding known as the path length effect (Parmentier et al., 2005; Guerard et al., 2009; Guerard and Tremblay, 2012). In addition, the path complexity, reflected in the number of path crossings (i.e., the number of times that a path between two successive locations crosses another path between two other locations), has been found to have an impact on performance as well. Memory accuracy was reduced as the number of path crossings increased (Kemps, 2001; Parmentier et al., 2005). Note that path characteristics have a temporal dimension as well as a spatial one. Nevertheless, the temporal order is determined by spatial considerations of proximity and complexity, and path characteristics have a direct impact on spatial WM performance, even when participants are probed on one location. Studies that found enhanced memory performance for structured spatial sequences explained the performance benefits in terms of grouping. Locations in structured spatial sequences were easily grouped into higher-order spatial configurations, whereas the locations in spatially unstructured sequences often disrupted grouping and consequently memory performance (Parmentier et al., 2005; Guerard et al., 2009).

We recently used these characteristics to evaluate the spatial configurations participants constructed in a spatial SI WM task (Magen and Emmanouil, 2018; see also Milchgrub and Magen, 2018). Participants in this task selected the spatial locations they memorized from an array of locations that were distributed randomly across the display. Performance in the spatial SI WM task was compared to a non-SI task, in which participants memorized random spatial sequences that were provided to them. The results revealed that relative to random sequences, the constructed SI spatial sequences had a shorter average path length, consisted of fewer path crossings, and followed more frequently simple and linear shapes. The structured SI representations demonstrated that participants had access

to metacognitive knowledge on the benefits of structure in spatial WM. Analysis of encoding RT showed that constructing these sequences involved planning and demanded resources. RT for the first location in the selected spatial sequence increased relative to subsequent locations and increased further with the sequence length. This pattern was absent from the non-SI condition, in which participants encoded provided spatial sequences. Finally, accuracy in the SI condition was higher than in the non-SI condition, even when the structure of the SI and non-SI spatial sequences was matched, demonstrating that self-initiation benefited performance beyond the benefit of structure.

The characteristics of spatially structured SI memory representations identified by Magen and Emmanouil (2018) were used to evaluate spatial organization in the current study as well. In addition, at the end of the experiment, we asked participants to describe the strategies they used to select the memory displays (cf. Magen and Berger-Mandelbaum, 2018). We predicted that if participants intended to memorize the spatial configuration of the memory displays they constructed, they would form spatially structured sequences. Otherwise, selection of targets based solely on non-spatial strategies should result in unstructured random spatial configurations.

## Methods

### Participants

Twenty-four students from the Hebrew University participated in a 1-h session. They provided informed consent before participating in the study for course credit or payment. The study was approved by the Hebrew University IRB.

### Stimuli and Design

Participants sat in a dimly lit room at a distance of 100 cm from the display and rested their head on a chin rest. In the SI and non-SI encoding conditions, participants were presented with an array of 12 pictures of real world objects (each picture measuring  $1.16^\circ \times 1.16^\circ$  of visual angle, presented on a gray background) appearing jittered (up to  $0.19^\circ$  of visual angle) in random locations within an invisible  $6 \times 6$  matrix (measuring  $10.32^\circ \times 10.32^\circ$  of visual angle). A pool of 504 pictures of real world objects from different categories (e.g., fruits, furniture, toys, and household objects) was used in the study for the SI and non-SI encoding conditions. Pictures were selected randomly on each trial, with the limitation that each picture could appear only once in each experimental block. A large set of targets was used such that different combinations of pictures would appear on each trial for each participant and would not direct participants to prefer certain strategies over others. For the same reason, the pictures were not controlled for low level features. Of the 12 presented pictures, participants memorized 1–7 pictures, selected by them in the SI condition, or selected randomly by the computer and provided to them in the non-SI condition. The set size manipulation was blocked, and block order was randomized across participants.

A single probe appeared at the center of the screen in both the SI and non-SI conditions. In the match condition (50% of the trials), the probe matched one of the selected targets, with equal frequency for targets at each serial position of the sequences selected during encoding. For instance, when three targets were selected and memorized, one third of the matched probes matched the target selected first, one third matched the target selected second, and the remaining trials matched the target selected third. In the non-match trials (remaining 50% of the trials), the probe was one of the targets in the original array that were not selected on that trial.

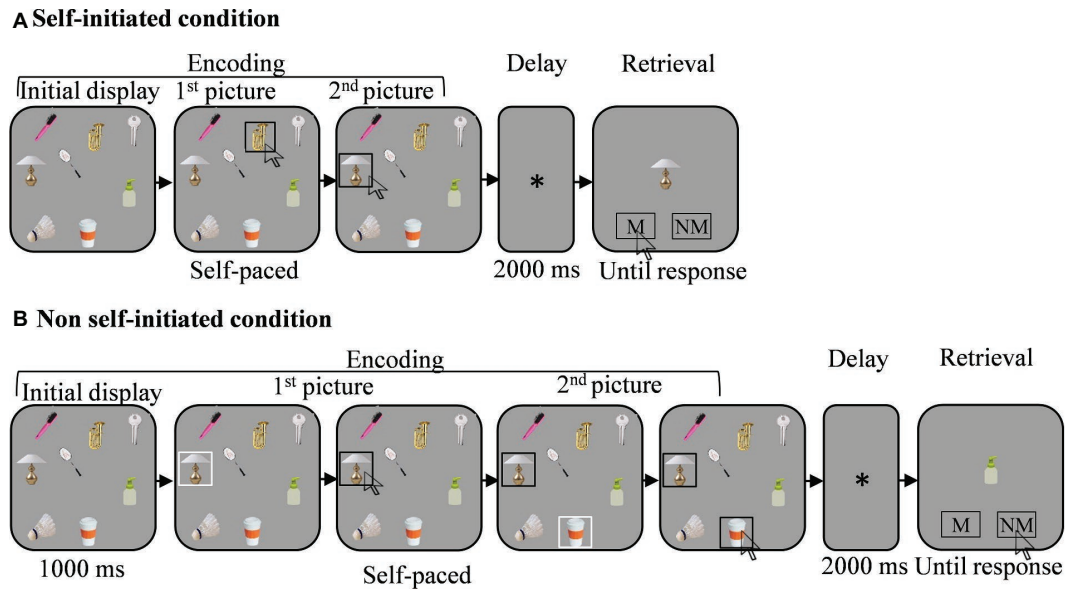
### Procedure

In both the SI and non-SI conditions, the trial started with the onset of a fixation cross for 500 ms, followed by the appearance of the target display (see **Figure 1**). In the SI condition, participants selected 1–7 targets sequentially, by clicking each selected target with the left key of the computer mouse. Following each selection, a black frame appeared around the selected target marking its selection. Encoding was self-paced.

In the non-SI condition, participants memorized 1–7 targets that were provided to them. In each trial, a white frame appeared around the 1–7 targets, one at a time, marking the target the computer selected. To equate the motoric response to the SI condition, the trial proceeded only after participants clicked each marked target in sequence. In response, the white frame turned black marking the selected target. Consequently, similar to the SI condition, participants controlled the pace and the duration of the encoding phase in the non-SI condition as well. In the non-SI condition, the first target in the sequence appeared 1,000 ms after the onset of the initial display. This was done to allow participants to scan the targets in the display before the first item was selected by the computer, as we hypothesized they would do in the SI encoding condition. In addition, a 200 ms delay was introduced in the non-SI condition between the selection (i.e., click) of the current target and the appearance of the white frame which marked the next to-be memorized target in the sequence.

In both the SI and the non-SI conditions, all chosen locations remained visible until the end of the encoding phase and disappeared 200 ms after the last location was clicked. The selected items remained visible throughout the trial to equate the conditions in the SI and non-SI conditions. We assumed that participants would plan their selections in the SI condition and therefore would hold the entire display in mind during encoding. Keeping the entire display visible during encoding in the SI and non-SI conditions reduced this potential advantage of self-initiation.

The maintenance and retrieval phases were identical in the SI and the non-SI conditions. Encoding was followed by a delay phase of 2,000 ms during which a fixation cross appeared at the screen center and which in turn was followed by the appearance of the central probe. The probe either matched or did not match one of the selected targets. Responses were registered by clicking with the mouse on one of two gray rectangles (measuring  $0.69^\circ \times 1.72^\circ$  of visual angle in height and width) with the words match and non-match written on them in Hebrew. The rectangles



**FIGURE 1 |** Illustration of the memory task (see text for further detail). Each trial began with the presentation of an initial spatial array of 12 pictures (for presentation purposes only eight pictures appear in the figure). Note that the pictures are not to scale). In the self-initiated (SI) condition (**A**) participants clicked sequentially on the pictures they wished to memorize. In the non-SI condition (**B**) participants clicked in sequence on the pictures selected randomly by the computer, which were marked by a white frame. Encoding was self-paced in both encoding conditions. In both conditions, encoding was followed by a short delay of 2,000 ms, after which a single probe appeared in the screen center, and either matched or did not match one of the selected pictures. Participants responded by clicking on one of the response areas that appeared with the probe. M, match. NM, non-match.

were presented at the bottom of the screen ( $3.72^\circ$  of visual angle below fixation) along with the probe. Accuracy was stressed in both the SI and non-SI conditions.

In each encoding condition, there were seven blocks of approximately 12 trials, one in each set size. There were 12 trials in set sizes 1, 2, 3, and 6, 16 trials in set size 4, 10 trials in set size 5, and 14 trials in set size 7. The number of trials varied between set sizes, because the probes in the match condition (half of the trials in each set size) appeared equally often in each of the serial order positions in the memory array.

Each participant performed the SI and the non-SI tasks in two separate blocks within the same session, with task order counterbalanced across participants. Each block (SI or non-SI) began with four short practice blocks consisting of two trials each, for set sizes 1–4. Throughout the experiment, an error message was presented on the screen for 500 ms following an incorrect response. The intertrial interval (ITI) was 1,500 ms in correct and error trials.

### Strategy Questionnaire

At the end of the experiment, participants filled out a questionnaire with two open-ended questions. The first question asked them to detail the strategies they used to select the targets in the SI condition, while the second question asked whether and how these strategies benefited memory performance.

## Results

Our analysis focused on three aspects of the results, encoding RT, the strategies participants reported using in selecting the

targets they memorized and the characteristics of the spatial configurations they constructed. An additional analysis focused on accuracy.

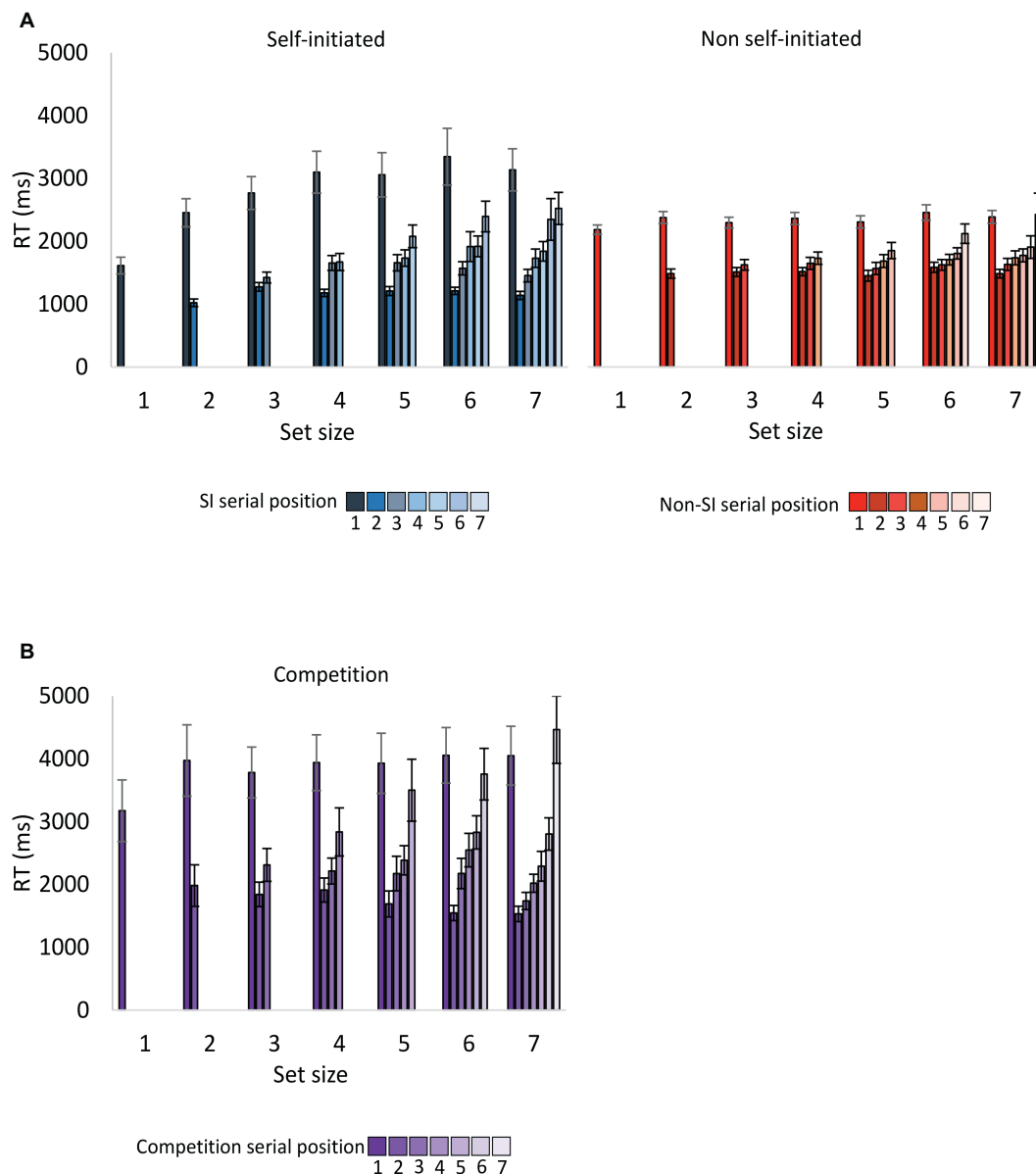
### Encoding Reaction Time

RTs for the entire sequence are presented in **Figure 2**. Given our previous studies (Magen and Emmanouil, 2018; Milchgrub and Magen, 2018), the analysis of encoding RT focused on RT for the first target in the sequence, which reflected planning. We also analyzed RT for the last target in the sequence, which showed how long participants took to review the target display before it disappeared, which reflected the participants perceived complexity of the array. In our previous studies, the non-SI condition yielded longer RTs for the last target in the sequence.

#### Reaction Time for the First Target in the Sequence

A repeated-measures ANOVA was conducted with the within-participant factors of encoding condition (SI and non-SI) and set size (1–7). The main effect of encoding was not significant  $F(1,23) = 3.59, p = 0.07, \eta_p^2 = 0.135$ , whereas the main effect of set size was significant  $F(6,138) = 10.58, p < 0.001, \eta_p^2 = 0.315$ . The interaction between set size and encoding was significant as well  $F(6,138) = 6.49, p < 0.001, \eta_p^2 = 0.220$ . Follow-up ANOVAs for each encoding condition revealed a significant set size effect in the SI condition  $F(6,138) = 9.45, p < 0.001, \eta_p^2 = 0.291$  (explained by linear and quadratic contrasts,  $F_s(1,23) = 19.51, 11.45, p_s < 0.01, \eta_p^2 = 0.459, 0.332$ , respectively), and a non-significant effect in the non-SI condition





**FIGURE 2 |** Encoding RT as a function of encoding condition, serial position and set size for **(A)** Experiment 1 **(B)** Experiment 2. Except for the first target, encoding RT for each target was calculated with respect to the selection of the previous target in the sequence. Error bars represent standard error of the mean. SI, self-initiated.

$F(6,138) = 1.53, p = 0.17, \eta_p^2 = 0.062$ . Thus, in the SI condition, RT for the first target increased with set size (up to set size 4), which suggests that participants planned the sequence of targets they selected before its execution and therefore required more time as set size increased.

#### Reaction Time for the Last Target in the Sequence (Set Sizes 2–7)

RT for the last target in the sequence was calculated with respect to the selection of the previous target in the sequence. The results yielded a non-significant main effect of encoding

condition  $F(1,23) = 0.09, p = 0.77, \eta_p^2 = 0.004$ . The main effect of set size was significant  $F(5,115) = 15.21, p < 0.001, \eta_p^2 = 0.398$ , and interacted significantly with encoding  $F(5,115) = 2.99, p < 0.05, \eta_p^2 = 0.115$ . Follow-up ANOVAs for each encoding condition revealed significant set size effects in the SI and non-SI conditions,  $F_s(5,115) = 19.45, 5.19, p_s < 0.001, \eta_p^2 = 0.458, 0.184$ , respectively. The interaction between encoding and set size reflected the findings that in the lower set sizes, RT for the last target in the sequence was longer in the non-SI condition, whereas in the higher set sizes, RT was longer in the SI condition.

## Encoding Strategies

The analysis of encoding strategies focused on the participants' responses to the strategy questionnaire and on the main characteristics of the spatial configurations they constructed.

### Strategy Questionnaire

The type of strategies and the number of participants who selected each of these strategies are presented in **Table 1**.

Participants used on average 1.92 strategies ( $SD = 0.78$ , Mode = 2, Range 1–4). Overall, five different strategies guided participants in their selections. Most frequently, participants selected targets based on semantic categories or on visual features (e.g., targets with similar colors or shapes or targets with distinct visual features). Five participants reported that they selected targets that fitted a story they created and three participants selected targets with reference to themselves. Of the 24 participants, only three participants reported using spatial strategies, by selecting targets positioned in close spatial proximity.

In their responses to the second question in the questionnaire, namely whether the strategy they used facilitated memory performance, all participants responded positively. They explained that the strategies they employed reduced the amount of information they memorized and helped identify matched and non-matched probes (by excluding targets that did not fit the semantic or visual regularities they set, or the story they created). Several participants explained specifically that it was easier to memorize visually distinct targets.

### Spatial Structure Analysis

Our next analysis focused on the characteristics of the spatial configurations that participants constructed, comparing them to the random computer-generated configurations. First, two characteristics of the path were analyzed, the average path length and the number of path crossings. An additional analysis examined the size of the spatial configuration. The results are presented in **Figure 3**.

We also explored the direction of movements participants made when they selected the sequence of targets during encoding, to reveal the overall shape of the spatial sequence, and whether participants tended to initiate the spatial sequence at the top

and left side as shown in previous studies (Magen and Emmanouil, 2018; Milchgrub and Magen, 2018). The results are presented in **Table 2**.

### Path Length (Set Sizes 2–7)

The average path length was significantly shorter in the SI condition as indicated by a main effect of encoding  $F(1,23) = 87.34$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.792$ . The main effect of set size and its interaction with encoding were non-significant  $F(5,115) = 1.01$ ,  $1.48$ ,  $ps = 0.41$ ,  $0.20$ ,  $\eta_p^2 = 0.042$ ,  $0.060$ , respectively.

### Path Crossings (Set Sizes 4–7)

The number of path crossings was significantly smaller in the SI condition relative to the non-SI condition  $F(1,23) = 34.64$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.602$ . The main effect of set size was also significant  $F(3,69) = 146.86$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.865$ , as was its interaction with encoding  $F(3,69) = 6.27$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.214$ . Follow-up ANOVAs showed that the effect of set size was significant in both the SI and the non-SI conditions,  $F_s(3,69) = 50.42$ ,  $158.68$ ,  $ps < 0.001$ ,  $\eta_p^2 = 0.687$ ,  $0.873$ , respectively. The interaction between encoding and set size resulted from a larger increase in the number of path crossings with set size in the non-SI condition.

### Configuration Size (Set Sizes 2–7)

In addition to path characteristics, we included a measure of the overall size of the spatial configuration. Very different paths (in terms of length and the number of crossings) can be formed between locations in the very same spatial configuration, and therefore these characteristics by themselves do not indicate whether participants selected targets in close proximity, relative to the random displays. The overall size of the spatial configurations was determined by calculating the centroid between all the locations in the configuration and then measuring the distance between each of the locations to the centroid. The analysis was based on the average distance between the centroid and all the targets in the configuration.

The average overall size of the spatial configurations was significantly smaller in the SI condition as reflected in a main effect of encoding  $F(1,23) = 47.87$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.675$ . The main effect of set size was also significant  $F(5,115) = 89.31$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.795$ , as the size of the configurations increased with set size in both encoding conditions. The interaction of set size with encoding was also significant  $F(5,115) = 8.95$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.280$ , as the difference between the two encoding conditions decreased as set size increased, probably due to a ceiling effect on the overall size of the configuration as the number of locations in the sequence increased. Follow-up ANOVAs showed that the set size effect was significant in the SI  $F(5,115) = 60.29$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.724$  and in the non-SI conditions  $F(5,115) = 49.36$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.682$ .

### Direction of Movements

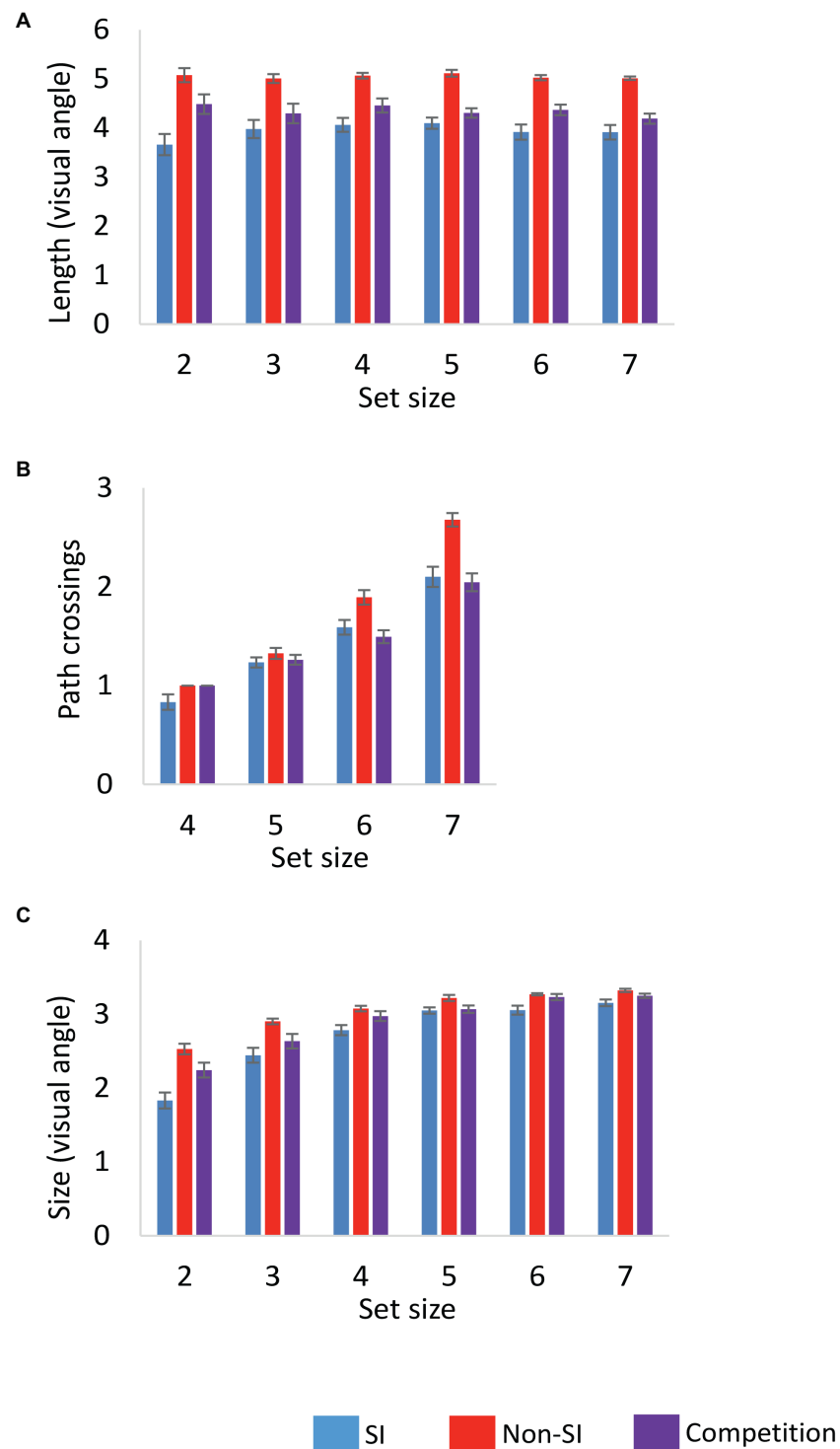
The final analysis examined the overall shape of the self-initiated spatial sequences, by examining the direction of movements

**TABLE 1 |** The number of participants who selected the different encoding strategies in Experiments 1 and 2.

Strategies	Visual	Semantic	Story	Self-reference	Spatial
<b>Experiment 1</b>					
SI	14	17	5	7	3
<b>Experiment 2</b>					
Competition	14*	8*	1*	3	0
Competition-SI	6	5	5	1	0

The results of Experiment 2 are presented separately for participants who selected strategies to disrupt memory performance (competition,  $n = 16$ ) and those who reported selecting targets to enhance memory performance (competition-SI,  $n = 8$ ) (see text for further details).

SI, self-initiated. \*Used these strategies in the opposite way, to disrupt grouping (see the "Results" section of Experiment 2, for further details).



**FIGURE 3 |** Characteristics of the spatial configurations as a function of encoding condition and set size in Experiments 1 and 2: **(A)** the mean path length, **(B)** the number of path crossings, and **(C)** configuration size. Error bars represent standard error of the mean. SI, self-initiated.

in the horizontal and vertical axes. The movements were also analyzed in the non-SI condition for comparison. Each movement was scored on both the horizontal and the vertical axes. On the horizontal axis, movements were divided into left, right,

and no horizontal movements (i.e., straight vertical movements), and in the vertical axis, movements were divided into down, up, and no vertical movements (i.e., straight horizontal movements). The analyses were motivated by the results of

**TABLE 2 |** Percent of movement directions in the construction of the spatial sequences in Experiments 1 and 2, for each encoding condition and set size, separately for the horizontal and vertical axes.

Set size	Condition	Horizontal			Vertical		
		Left	Right	None	Down	Up	None
Experiment 1							
2	SI	37.85	42.71	19.44	33.68	40.63	25.69
	Non-SI	37.15	46.18	16.67	42.01	43.06	14.93
3	SI	38.89	40.80	20.31	42.01	35.76	22.22
	Non-SI	45.14	39.93	14.93	44.62	41.84	13.54
4	SI	41.06	40.45	18.49	39.84	37.24	22.92
	Non-SI	40.89	43.40	15.71	42.45	42.97	14.58
5	SI	39.90	42.08	18.02	41.56	35.73	22.71
	Non-SI	43.65	43.33	13.02	44.79	42.50	12.71
6	SI	38.82	42.50	18.68	40.28	35.83	23.89
	Non-SI	43.19	42.29	14.51	41.88	42.71	15.42
7	SI	39.09	40.48	20.44	41.17	35.86	22.97
	Non-SI	43.40	42.06	14.53	42.81	43.40	13.79
Average SI		39.27	41.50	19.23	39.76	36.84	23.40
Average non-SI		42.24	42.87	14.90	43.09	42.75	14.16
Experiment 2							
2	Competition	44.44	40.63	14.93	38.89	38.54	22.57
3	Competition	37.67	44.79	17.53	40.28	38.54	21.18
4	Competition	37.33	44.01	18.66	42.27	38.45	19.27
5	Competition	40.31	42.92	16.77	42.40	38.65	18.96
6	Competition	40.07	43.47	16.46	41.11	39.79	19.10
7	Competition	39.68	42.76	17.56	41.67	38.24	20.09
Average competition		39.92	43.10	16.99	41.10	38.70	20.20

SI, self-initiated.

our previous studies (Magen and Emmanouil, 2018; Milchgrub and Magen, 2018). First, the number of straight vertical and horizontal movements were compared between the SI and non-SI conditions. Second, to examine whether movements were initiated at the top and left side, we created a difference score for each axis and compared these scores across the two encoding conditions. As can be seen in **Table 2**, the pattern of movements in the SI spatial sequences deviated from the random movements generated in the non-SI condition. Movements in the same horizontal or vertical axis (creating straight lines) were more frequent in the SI condition. An ANOVA with encoding and set size revealed a main effect of encoding,  $F(1,23) = 66.59$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.743$  showing a larger percent of straight horizontal movements in the SI condition. The main effect of set size and its interaction with encoding were not significant  $F_s < 1$ . The same pattern was observed for the straight vertical movements, showing a significant main effect of encoding  $F(1,5) = 21.56$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.484$ , whereas the main effect of set size and its interaction with encoding were not significant  $F_s < 1$ . Thus, the SI configurations consisted of more linear shapes.

Two additional ANOVAs explored whether vertical (more downward than upward movements) or horizontal (more rightward than leftward movements) biases were present in the SI configurations, by comparing the difference scores between the SI and non-SI conditions. The two ANOVAs yielded null effects. The vertical bias yielded non-significant effects of encoding, set size, and a non-significant interaction between them, all

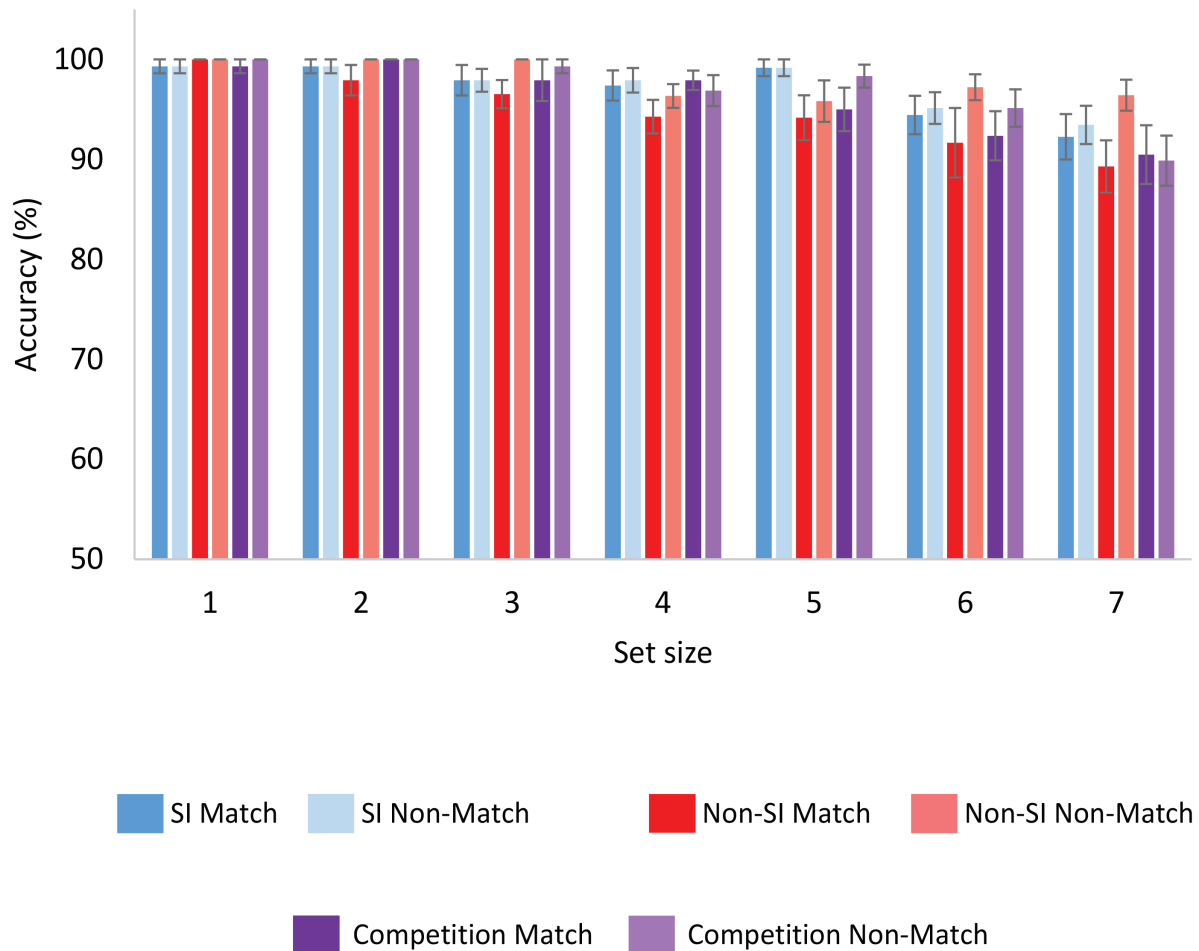
$F_s < 1.54$ ,  $p_s > 0.18$ ,  $\eta_p^2 < 0.063$ . The ANOVA examining the horizontal bias revealed a non-significant main effect of encoding  $F < 1$ . The main effect of set size was significant, mainly due to random variations in movements in the non-SI condition (see **Table 2**),  $F(1,5) = 2.49$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.098$ . The interaction between encoding and set size was non-significant  $F < 1$ .

### Accuracy

Accuracy in the match and non-match trials was compared across the SI and non-SI encoding conditions. Accuracy in the match condition was averaged across all serial positions in each set size, to obtain a single measure of accuracy. As evident from **Figure 4**, although accuracy was almost at ceiling, there were small differences between the different conditions. A repeated-measured ANOVA with encoding, set size, and probe condition (match or non-match) as within-participant factors showed non-significant main effects of encoding and probe  $F(1,23) = 2.31$ ,  $p = 0.14$ ,  $\eta_p^2 = 0.091$  and  $F(1,23) = 3.42$ ,  $p = 0.08$ ,  $\eta_p^2 = 0.129$ , respectively. Nevertheless, the interaction between them was significant,  $F(1,23) = 5.76$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.200$ . The main effect of set size was significant  $F(6,138) = 10.57$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.315$  and did not interact significantly with encoding  $F(6,138) = 1.58$ ,  $p = 0.16$ ,  $\eta_p^2 = 0.064$ , or probe  $F(6,138) = 1.08$ ,  $p = 0.38$ ,  $\eta_p^2 = 0.045$ . The three-way interaction of encoding, probe, and set size was non-significant  $F(6,138) < 1$ .

To follow-up on the interaction between the encoding and probe factors, two additional ANOVAs were conducted separately for the match and non-match trials. The analysis showed a





**FIGURE 4 |** Accuracy as a function of encoding condition, set size and probe condition (match or non-match) for Experiments 1 and 2. Error bars represent standard error of the mean. SI, self-initiated.

significant main effect of encoding in the match trials  $F(1,23) = 4.86, p < 0.05, \eta_p^2 = 0.174$ , but not in the non-match trials  $F < 1$ . Thus, accuracy was significantly higher in the SI condition, but only in match trials when a target from the memory display appeared as the probe.

## Discussion

The results of Experiment 1 demonstrated that the spatial sequences participants constructed were spatially structured relative to the non-SI sequences, although space was task irrelevant. Relative to the non-SI condition, the memory displays constructed in the SI condition were smaller in size and were characterized by a shorter average path length, by a smaller number of path crossings and by more linear shapes. We assume that participants imposed the spatial structure on the memory representations they constructed, to benefit memory performance, rather than, for instance, ease the selection process during encoding. We assume that structure was intended to benefit memory for several reasons. First, encoding was self-paced, and participants could spend as much time as they needed

to familiarize themselves with each target. It seems unlikely therefore, that they would impose a spatial structure during encoding only for selection purposes, structure that could disrupt the implementation of the non-spatial encoding strategies. Most importantly, the spatial structure that participants imposed in this study is consistent with the structure that has been shown in previous non-SI and SI studies to benefit memory performance (e.g., Kemps, 2001; Parmentier et al., 2005; Magen and Emmanouil, 2018). Consistent with our previous studies (Magen and Berger-Mandelbaum, 2018; Magen and Emmanouil, 2018; Milchgrub and Magen, 2018), the constructed configurations were the result of effortful planning as reflected by encoding RT for the first target in the sequence, which increased with set size only in the SI condition.

Accuracy was high in both encoding conditions demonstrating that participants adjusted encoding RT to maximize memory performance in both conditions. Nevertheless, accuracy was still significantly higher in the SI condition when the probe matched one of the memorized items. The high and similar accuracy levels could explain why, unlike our previous studies,

RT for the last target in the sequence was similar in the two encoding conditions. Perhaps the nature of the stimuli in the task, in addition to the unlimited encoding time, allowed participants to encode items in the SI and the non-SI conditions such that the perceived difficulty of the maintained memory representations was similar across the two encoding conditions.

Participants' reports on the strategies they used revealed that the most frequently reported strategies were non-spatial, grouping items based on semantic categories or visual features. Spatial strategies were scarcely reported, which suggests that participants were largely unaware of the spatial characteristics of the memory representations they formed. Remember that the visual targets were distributed randomly across the display on each trial, and therefore it is unlikely that the visual targets that fitted the participants' non-spatial strategies were consistently placed in close proximity. Thus, the construction of the spatially organized configurations must have constrained the implementation of the non-spatial strategies to some degree.

The non-spatial visual strategies used to group the memory targets were similar to the strategies identified in our previous study (Magen and Berger-Mandelbaum, 2018), although in that study participants based their selections of real-world objects more frequently on visual features than on semantic categories. The limited set size conditions of 3 or 4 items in our previous study could explain this difference. It is possible that participants have found it necessary to employ diverse strategies in this study, as set size increased to seven items. This suggestion is supported by the finding that the strategies of self-reference and the creation of a story were only used in the current study and not in the study of Magen and Berger-Mandelbaum (2018).

Taken together, the results of Experiment 1 suggest that, similar to our previous studies in which space was task relevant, participants have metacognitive knowledge on the role of space in VWM. This knowledge was nevertheless mostly implicit, as only a handful of participants reported spatial proximity as one of the strategies they used to construct their memory representations. Note that in our previous study (Magen and Berger-Mandelbaum, 2018), when space was relevant to task performance, participants provided verbal descriptions of the spatial strategies they used to place the targets on the circular array they were presented with during encoding. Thus, spatial strategies by themselves could be explicit in other tasks.

Experiment 2 utilized a non-verbal manipulation to uncover whether these spatial strategies could be manipulated flexibly, although the metacognitive knowledge that guided them was implicit.

## EXPERIMENT 2

The main purpose of Experiment 2 was to explore whether participants would construct less structured spatial configurations under opposite instructions. Participants in this task were asked to construct memory displays for a hypothetical competitor in a memory contest, displays that in effect would disrupt memory performance. We had used this manipulation in the past to reveal the flexibility of spatial SI memory representations (Magen and

Emmanouil, 2018). The results of this experiment showed that when asked to disrupt memory performance, participants constructed spatial sequences that were characterized by a longer average path length, a larger number of path crossings, and by more non-linear shapes, relative to a SI encoding condition. Thus, in the "competition" task, participants manipulated the same characteristics that they used to construct structured SI representations, revealing flexible use of the metacognitive knowledge on the impact of these characteristics on memory performance.

Experiment 2 used the "competition" manipulation with the SI VWM task of Experiment 1, to examine whether participants would manipulate the spatial characteristics of the visual memory representations they would construct for a hypothetical competitor. Similar to Experiment 1, we also asked participants to provide verbal reports on the strategies that guided the selection of the targets they memorized.

## Methods

### Participants

A new group of 24 students from the Hebrew University participated in a 30-min session for course credit or payment. Participants provided informed consent before participating in the study. The study was approved by the Hebrew University IRB.

The task was identical to the SI condition in Experiment 1, except for task instructions, which asked participants to select locations for a hypothetical competitor in a memory contest. As in Experiment 1, participants filled out a strategy questionnaire at the end of the experiment, in which they were asked to detail the strategies they used to select the targets during encoding. In Experiment 2, the participants were asked in addition whether and how the strategies they selected disrupted memory performance.

## Results

The main analysis focused on comparing the results of the competition condition in Experiment 2, to each of the SI and non-SI conditions in Experiment 1, in terms of encoding RT, the spatial and non-spatial encoding strategies they used, and accuracy.

As detailed in subsequent sections, the analysis of participants responses to the strategy questionnaire showed that 8 of the 24 participants reported using strategies with the attempt to enhance memory performance rather than disrupt it. These participants were not removed *post hoc* from the main analysis; however, at the end of the "Results" section, we included an additional analysis that compared the results of these 8 participants to the remaining 16 participants who followed the competition instructions.

### Comparison of Experiments 1 and 2

#### Encoding Reaction Time

##### *Reaction Time for the First Target in the Sequence*

As evident in Figure 2 and confirmed in the analyses below, overall RT for the first target in the sequence was longer in the competition condition relative to the SI and the non-SI conditions. The first of two mixed-effects ANOVAs with set size as a within-participant factor and encoding condition

(competition and SI) as a between-participant condition revealed a significant main effect of encoding  $F(1,46) = 5.41, p < 0.05, \eta_p^2 = 0.105$ , a significant main effect of set size  $F(6,276) = 5.81, p < 0.001, \eta_p^2 = 0.112$ , and a non-significant interaction between them,  $F < 1$ . The second ANOVA that compared the competition and the non-SI conditions showed a main effect of encoding  $F(1,46) = 15.22, p < 0.001, \eta_p^2 = 0.249$ , a non-significant main effect of set size  $F(6,276) = 1.50, p = 0.18, \eta_p^2 = 0.031$ , and a non-significant interaction between them  $F < 1$ .

The results of Experiment 1 showed that RT for the first target in the sequence increased with set size in the SI condition but not in the non-SI condition. The results of the competition condition were similar to the non-SI condition, revealing a non-significant main effect of set size  $F(6,138) < 1$ . Nevertheless, when the competition and the SI conditions were compared, the two-way interaction of set size and encoding was non-significant. This is most likely due to the inclusion in the data of Experiment 2, the group of participants who adopted an SI encoding strategy rather than a “competition” strategy (see below).

#### *Reaction Time for the Last Target in the Sequence (Set Sizes 2–7)*

Comparing the competition and the SI conditions yielded a main effect of encoding  $F(1,46) = 14.22, p < 0.001, \eta_p^2 = 0.236$ , as RT for the last target in the sequence was longer in the competition condition relative to the SI condition. The main effect of set size was also significant  $F(5,230) = 22.45, p < 0.001, \eta_p^2 = 0.328$  and did not interact significantly with encoding  $F(5,230) = 1.44, p = 0.24, \eta_p^2 = 0.030$ . RT for the last item was longer in the competition condition relative to the non-SI condition as well, as shown by a main effect of encoding  $F(1,46) = 14.38, p < 0.001, \eta_p^2 = 0.238$ . The main effect of set size was also significant  $F(5,230) = 15.13, p < 0.001, \eta_p^2 = 0.248$  and interacted significantly with encoding  $F(5,230) = 3.39, p < 0.05, \eta_p^2 = 0.069$ , as the difference between the two conditions increased with set size.

## Encoding Strategies

### *Strategy Questionnaire*

The results of the strategy questionnaire are presented in **Table 1**. The results of the 8 participants who reported using strategies to enhance memory performance are presented separately from the results of the remaining 16 participants.

Overall, participants in Experiment 2 used on average 1.79 strategies ( $SD = 0.59$ , Mode = 2, Range 1–3). As shown in **Table 1**, participants who reported using encoding strategies to disrupt memory performance used similar strategies to the participants in the SI condition in Experiment 1, but in the opposite way. Specifically, they selected targets that were visually dissimilar, and non-distinct, and selected targets from different semantic categories. Participants who created a story reported that they associated targets to the story in a personal way that would be difficult for others to decipher. Finally, several participants selected targets related to themselves, assuming that it would not benefit the memory of others. None of the participants mentioned space as a strategy for target selection.

Participants explained that the targets they selected disrupted memory performance because they were not easily associated with each other and therefore increased the load on memory.

The eight participants who selected targets to enhance memory performance used the same strategies and explanations as in the SI condition in Experiment 1 (see **Table 1**).

### *Spatial Structure Analysis*

Overall, participants in Experiment 2 constructed memory representations that were less spatially structured compared to the SI condition in Experiment 1. The results are presented in **Figure 3**.

#### *Path Length (Set Sizes 2–7)*

Comparing the competition and the SI condition showed a significant main effect of encoding  $F(1,46) = 6.89, p < 0.05, \eta_p^2 = 0.130$ , as the average path length was longer in the competition condition. The main effect of set size and its interaction with encoding were not significant,  $F < 1$ , and  $F(5,230) = 1.69, p = 0.14, \eta_p^2 = 0.035$ , respectively. Comparing the competition to the non-SI condition also showed a significant main effect of encoding  $F(1,46) = 40.76, p < 0.001, \eta_p^2 = 0.470$ . Although the path length in the competition condition was longer than in the SI condition, it was still significantly shorter than in the non-SI condition. The main effect of set size and its interaction with encoding were not significant  $F_s < 1$ .

#### *Path Crossings (Set Sizes 4–7)*

In contrast to the path length, the average number of path crossings was similar in the competition and SI conditions. The main effect of encoding was non-significant  $F < 1$ , while the main effect of set size was significant  $F(3,138) = 100.36, p < 0.001, \eta_p^2 = 0.686$ . The two factors did not interact significantly  $F(3,138) = 1.40, p = 0.25, \eta_p^2 = 0.030$ . The average number of path crossings was significantly smaller in the competition condition relative to the non-SI condition  $F(1,46) = 47.79, p < 0.001, \eta_p^2 = 0.510$ . The main effect of set size was significant as well  $F(3,138) = 191.76, p < 0.001, \eta_p^2 = 0.807$ , as was its interaction with encoding  $F(3,138) = 12.12, p < 0.001, \eta_p^2 = 0.209$ . The interaction reflected the observation that the difference between the competition and non-SI conditions increased with set size.

#### *Configuration Size (Set Sizes 2–7)*

The comparison of the configuration size between the competition and the SI conditions yielded a significant main effect of encoding  $F(1,46) = 7.91, p < 0.01, \eta_p^2 = 0.147$ , showing that participants constructed larger spatial configurations in the competition condition. The main effect of set size was significant as well  $F(5,230) = 104.28, p < 0.001, \eta_p^2 = 0.694$ , and its interaction with encoding was marginally significant  $F(5,230) = 2.27, p = 0.05, \eta_p^2 = 0.047$ . The difference between the two conditions decreased with set size, most likely due to a ceiling effect on the overall size of the spatial configurations. The competition condition differed significantly from the non-SI condition as well, as the spatial configurations in the competition condition were on average smaller

than in the non-SI condition  $F(1,46) = 10.35$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.184$ . The main effect of set size was significant  $F(5,230) = 90.49$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.663$ , and it did not interact significantly with encoding  $F(5,230) = 2.03$ ,  $p = 0.08$ ,  $\eta_p^2 = 0.042$ .

#### Path Shape

We focus in this analysis only the percent straight vertical and horizontal movements, which showed differences between the SI and non-SI conditions in Experiment 1. As shown in **Table 2**, the percent of linear shapes in the competition condition was intermediate between the SI and non-SI conditions. Straight horizontal and vertical movements were more frequent in the SI condition than in the competition condition, as the main effect of encoding was significant in both ANOVAs,  $F(1,46) = 4.51$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.089$  and  $F(1,46) = 5.06$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.099$ , respectively. The main effects of set size or their interaction with encoding were not significant in either ANOVA, all  $F$ s  $< 1.13$ . Comparison of the competition and non-SI conditions revealed more frequent straight horizontal and vertical movements in the non-SI condition, as reflected in the main effects of encoding in both ANOVAs,  $F(1,46) = 40.94$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.471$  and  $F(1,46) = 5.10$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.100$ , respectively. The main effects of set size or their interaction with encoding were non-significant, all  $F$ s  $< 1.02$ .

#### Accuracy

Accuracy was compared between the competition condition and the SI and non-SI conditions, although accuracy in the competition condition may be more difficult to interpret, as participants memorized information that they selected with the intent to disrupt memory performance. Similar to Experiment 1, accuracy was high in the competition condition as well, (see **Figure 4**). The analyses showed that accuracy was the same in the competition and SI conditions, as the main effect of encoding was non-significant  $F < 1$ , and neither was the main effect of probe (match or non-match)  $F(1,46) = 1.12$ ,  $p = 0.30$ ,  $\eta_p^2 = 0.024$ . Only the main effect of set size was significant  $F(6,276) = 16.25$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.261$ . None of the interactions between the three factors were significant, all  $F$ s  $< 1$ . The competition and the non-SI condition yielded similar accuracy levels as well,  $F < 1$  for the main effect of encoding. The other two main effects of set size and probe were significant  $F(6,276) = 16.19$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.260$ , and  $F(1,46) = 5.97$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.115$ , respectively. None of the interactions were significant, all  $F$ s  $< 1.75$ ,  $p$ s  $> 0.19$ .

#### Comparing Participants Based on Reported Strategies

In this section, we compared the data of the 8 participants who reported using encoding strategies to enhance memory performance to the remaining 16 participants who reported using strategies to disrupt performance. Because of the small number of participants in each group, we averaged the data across set sizes to obtain one measure of RT, structure, and accuracy from each participant. Except for accuracy, which was evaluated in a mixed-model ANOVA, the two groups of

participants were compared by independent  $t$ -tests (all two-tailed). Figures depicting the full set of results appear in the **Supplementary Figures SA1–SA3**.

#### Reaction Time for the First Target in the Sequence

RTs for the first target in the sequence were longer among the participants who followed the competition instructions relative to the participants who selected memory displays to enhance memory performance ( $M = 4275.19$ ,  $SD = 1868.41$  and  $M = 2980.05$ ,  $SD = 1545.51$ , respectively). This difference however was not significant with a two-tailed test,  $t(22) = 1.69$ ,  $p = 0.11$ , Cohen's  $d = 0.76$ . As can be seen in **Supplementary Figure SA1**, RT for the first item increased with set size only in the group of participants who selected targets to enhance memory performance.

#### Reaction Time for the Last Target in the Sequence (Averaged Across Set Sizes 2–7)

RTs for the last target were longer in the group of participants who followed the competition instructions relative to participants who did not follow these instructions ( $M = 3606.09$ ,  $SD = 1688.14$  and  $M = 2214.71$ ,  $SD = 574.48$ , respectively). The difference was significant  $t(22) = 2.25$ ,  $p < 0.05$ , Cohen's  $d = 1.10$ .

#### Path Length (Averaged Across Set Sizes 2–7)

The average path length was significantly longer in the group of participants who followed the competition instructions ( $M = 4.54^\circ$ ,  $SD = 0.48$ ) compared to the group of participants who reported selecting memory representations to enhance performance ( $M = 3.97^\circ$ ,  $SD = 0.33$ ),  $t(22) = 3.08$ ,  $p < 0.01$ , Cohen's  $d = 1.38$ .

#### Path Crossings (Averaged Across Set Sizes 4–7)

The number of path crossings was similar in the two groups of participants ( $M = 1.43$ ,  $SD = 0.13$ , and  $M = 1.49$ ,  $SD = 0.17$ , for the participants who followed the competition instructions and those who did not follow it, respectively),  $t(22) = -1.00$ ,  $p = 0.33$ , Cohen's  $d = -0.40$ .

#### Configuration Size (Averaged Across Set Sizes 2–7)

The size of the spatial configuration was significantly larger in the group of participants who followed the competition instructions ( $M = 2.99^\circ$ ,  $SD = 0.18$ ) relative to the group of participants who did not follow these instructions ( $M = 2.71^\circ$ ,  $SD = 0.16$ ),  $t(22) = 3.66$ ,  $p < 0.001$ , Cohen's  $d = 1.64$ .

#### Accuracy

Accuracy was evaluated by a mixed model ANOVA with probe type as a within participant factor and group as a between-participant factor. Accuracy in the two groups was similar and high overall. The results of the ANOVA yielded a non-significant main effect of group  $F(1,22) < 1$ , probe  $F(1,22) = 1.36$ ,  $p = 0.26$ ,  $\eta_p^2 = 0.058$ , and a non-significant interaction between them  $F(1,22) < 1$ .



## Discussion

The main question addressed in Experiment 2 was whether participants would construct less structured spatial sequences when asked to construct spatial sequences that would disrupt memory performance. Indeed, relative to the representations in the SI condition of Experiment 1, the memory representations constructed in the competition condition consisted of fewer linear shapes, were overall larger, and had a longer average path length. Participants' verbal reports focused exclusively on non-spatial strategies. The results of Experiment 2 further suggest that the spatial structure that participants imposed on the representations they constructed in the SI condition was intended to enhance memory rather than ease selection processes during encoding, as this structure was abolished when participants were asked to construct representations that would disrupt memory performance. Thus, participants in this experiment also had accesses to the metacognitive knowledge on the role of space in VWM. Although this knowledge was implicit (based on verbal reports), participants exerted control over the spatial strategies, by flexibly manipulating the spatial characteristics of the memory displays to disrupt memory performance.

The number of path crossings was the same in the competition and the SI conditions of Experiment 1, suggesting that participants identified proximity as the major factor that influenced memory performance. Alternatively, it is possible that participants have no accesses to the metacognitive knowledge on the impact of path crossings on performance. This explanation is less likely since participants in a previous spatial SI WM task did increase the number of path crossings in a competition condition compared to a SI condition (Magen and Emmanouil, 2018).

While less structured than in the SI condition, the memory representations in the competition condition were more structured than in the random non-SI condition, which suggests that the initial tendency is to construct structured spatial sequences that is manipulated and disrupted in the competition condition. Note that when the eight participants who did not follow the competition instructions were removed from the analysis, the difference between the competition and non-SI conditions remained significant in path length and the number of path crossings, but the overall size of spatial configurations was similar in these two encoding conditions.

Encoding RT for the last target in the sequence was longer in the competition condition relative to the SI and non-SI conditions. This finding confirms that participants in the competition condition constructed memory representations that they perceived to be more challenging. Nevertheless, accuracy was similar in the competition and SI conditions demonstrating that participants in Experiment 2 also adjusted encoding RT to reach almost ceiling performance.

Finally, 8 of the 24 participants in Experiment 2 did not follow the competition instructions. Comparing the results of the participants who attempted to enhance memory performance to those who attempted to disrupt memory performance showed clear differences in the spatial characteristics of the memory

displays they constructed. This dissociation within the results of Experiment 2 provided additional support for the overall results of this study, and especially for the direct association participants (implicitly) made between spatial structure and the ease or difficulty of memory performance.

## GENERAL DISCUSSION

Various situations in everyday life require individuals to shape the content of their memory representations. We have recently begun to explore this aspect of memory we termed SI memory. In the current study, we focused on the spatial structure of SI VWM representations in a memory task in which space was task irrelevant. The results of two experiments demonstrated that when asked to enhance memory performance, participants planned and constructed spatially structured memory representations that relative to random provided representations, were overall smaller, consisted of more linear shapes and the spatial sequence path was on average shorter and consisted of fewer path crossings. These spatial structures were mostly compatible with the results of our previous studies on spatial SI WM (Magen and Emmanouil, 2018; Milchgrub and Magen, 2018) and with the literature on structured provided spatial WM (e.g., Parmentier et al., 2005; Guerard et al., 2009). When asked to disrupt memory performance, participants constructed less spatially structured representations, which relative to the SI representations, were larger, with longer paths, and with fewer linear shapes. The number of path crossings, in contrast, was similar in the two experiments, which suggests that participants considered the size and shape of the overall spatial configuration as the important characteristics that influence memory performance in SI VWM. Because encoding was self-paced and participants were presented with a single central probe during retrieval, we speculate that the construction of the spatially structured configurations was intended to benefit maintenance.

Participants provided verbal reports on the strategies they used to construct the memory representations. Participants in the SI condition indicated that they most frequently selected targets that could be grouped by semantic categories or visual features. These very same non-spatial grouping strategies were abolished (i.e., used in the opposite way) in the competition condition when participants were asked to disrupt memory performance. The spatial aspects of the constructed memory representations were largely absent from participants' verbal reports, although spatial structure was clearly imposed on the memory representations. Moreover, the spatial structure likely interfered with the implementation of the non-spatial strategies, since it constrained the choices that could be made based on the non-spatial grouping cues.

An interesting aspect of the present study is the finding that participants invested time and resources in the construction of spatially structured representations (as demonstrated by encoding RT), but were largely unaware of it. While we cannot isolate encoding RT related to the selection of the visual and the spatial characteristics of the constructed memory

representations, our previous studies have shown that the construction of structured spatial representations by themselves was time consuming (Magen and Berger-Mandelbaum, 2018; Magen and Emmanouil, 2018). It would be interesting in future studies to tease apart the construction of the visual and spatial characteristics of the memory representations. These processes may operate independently as behavioral and neural evidence on provided VWM suggest that visual targets and the spatial configurations in which they are embedded in WM are dissociable (Ackerman and Courtney, 2012; Xie and Zhang, 2017).

In this and in our previous studies, we claim that SI WM representations are based on metacognitive knowledge on the basic structure of WM representations. As far as we know, the ability to apply this metacognitive knowledge in the construction of one's own memory representations remains an unexplored topic. The results of the current study suggest that this knowledge can guide strategy selection in a controlled way that has a profound influence on behavior, but nevertheless remain implicit. These results are consistent with findings and models from the research literature on metacognition, which had suggested that implicit metacognitive knowledge can guide strategy selection and implementation, and that controlled metacognitive processes can be guided by different degrees of metacognitive awareness (Cary and Reder, 2002). The results of Experiment 2 are even more intriguing in this regard, as they show that even the more sophisticated strategic controlled processes in the competition condition (i.e., manipulating structure in the opposite way) remain implicit.

Thus, the implementation of the spatial strategies in the construction of SI memory representations and their absence from the participants' verbal reports revealed different degrees of metacognitive awareness of this knowledge. We are unaware of previous studies that can illuminate this aspect of the results. However, some insights can be gained from studies in perception, which have shown dissociations in the estimated strength of different Gestalt organization cues when objective and subjective measures were used (Schmidt and Schmidt, 2013; Montoro et al., 2017). For instance, Schmidt and Schmidt (2013) showed that grouping by shape produced stronger effects on (objective) behavior than grouping by brightness, although both cues were judged to be equal in strength based on subjective ratings. The authors suggested that although the objective and subjective measures relied on the same visual input, the strength of each grouping cue was represented differently when it served the objective and subjective tasks. Unlike the studies just described, the tasks in this study were all self-initiated and subjective in nature, and organization was constructed by individuals rather than passively perceived. That said, the idea that organization in visual representations can be represented distinctly along different levels of the cognitive system is relevant and would be an interesting topic for future studies.

The memory representations constructed in the current study were spatially structured, consistent with our previous spatial SI WM studies (Magen and Emmanouil, 2018; Milchgrub and Magen, 2018). Nevertheless, several dissociations that were observed between this and the previous studies could point to fundamental differences between spatial and visual SI WM. First, the overall shape of the spatial structures constructed

in the current study was less organized than in the previous spatial SI WM studies, consisting of fewer linear shapes. Furthermore, in our previous spatial SI WM tasks, we observed a consistent bias to initiate the spatial sequence at the top left side, which did not occur in the current study. Thus, while the spatial structure was shown to be quite robust in this SI VWM task, participants probably first selected visual targets that matched their non-spatial strategies and then built around them a structured spatial configuration that matched the non-spatial rule as best as possible.

Further distinctions between spatial and visual SI WM were observed in the competition experiments. The number of path crossings was manipulated in the competition condition (i.e., increased relative to the SI condition) only in the spatial SI WM tasks. Thus, while in the tasks used in the current study participants minimized the number of path crossings to enhance performance, and therefore on some level acknowledged its potential impact on memory performance, this aspect of the spatial configuration was overlooked when participants attempted to disrupt memory performance. This dissociation between spatial and visual SI WM could suggest that the underlying mental models that guide the construction of the memory representations in these two types of tasks vary to some degree.

Differences in the underlying mental models or regularities in spatial and visual SI WM could also explain why RT for the first item in the sequence was longer in the competition condition relative to the SI condition in the current study, but not in our previous spatial SI WM studies (Magen and Emmanouil, 2018, see also Magen and Emmanouil, 2019). We speculate that structured spatial configurations may be governed by one set of regularities (e.g., path length, number of crossings) that could be implemented implicitly and could be easily abolished when required. Grouping of visual targets, on the other hand, could be based on several different types of regularities that participants employ explicitly, and that first need to be established before they are abolished. Furthermore, constructing representations to disrupt memory performance based on idiosyncratic regularities may require planning as well.

Finally, accuracy was almost at ceiling across all conditions. Nevertheless, the SI condition yielded significantly higher accuracy rates than the non-SI condition. This effect, however, was small and was restricted to the match trials, probably due to the overall high accuracy rates in this task. The advantage of the SI over the non-SI condition could be related to any number of factors, the structured nature of the SI representations, benefit from self-initiation, or both. In two previous studies on spatial SI WM, we found that accuracy was enhanced in the SI condition relative to the non-SI condition, even when they were matched in structure, demonstrating additive benefits of structure and self-initiation on performance (Magen and Emmanouil, 2018, 2019). Several processes may underlay this additional advantage of self-initiation. For instance, long-term memory performance is often enhanced for self-referenced or self-generated information (Slamecka and Graf, 1978; Cunningham et al., 2008). Furthermore, the control participants had over the memory displays in the SI condition may have increased their sense of agency, which

has been shown to have beneficial effects on memory performance (e.g., Murty et al., 2015).

While the benefits in memory performance associated with self-initiation are important and should be explored in future studies, the focus of the present study was on the underlying structure of SI VWM representations. These representations reveal the manner by which individuals shape their world for short or long durations, behavior that is prevalent and critical for efficient everyday functioning, but is largely absent from the WM literature. More generally, only a small number of studies have examined the manner by which individuals organize their surroundings in other domains. For instance, Solman and Kingstone (2017) examined how individuals organize objects during online performance of a novel task the authors created. The results showed that participants adopted strategies that led them to organize their space in accordance with task demands. Several of the strategies were associated with task performance. The cognitive processes that underlay the self-initiated behavior of individuals as they shape their environment should gain more attention in the research literature. Within this vast topic, we have begun to outline the basic structure of SI WM representations, which capture individuals' metacognitive knowledge regarding the structure of memory, the strengths and weaknesses of their cognitive system, and their adaptability to an ever-changing complex world.

## REFERENCES

- Ackerman, C. M., and Courtney, S. M. (2012). Spatial relations and spatial locations are dissociated within prefrontal and parietal cortex. *J. Neurophysiol.* 108, 2419–2429. doi: 10.1152/jn.01024.2011
- Berger-Mandelbaum, A., and Magen, H. (2019). Self-initiated object-location memory in young and older adults. *Aging Neuropsychol. Cognit.* 26, 58–85. doi: 10.1080/13825585.2017.1399981
- Bor, D., Duncan, J., Wiseman, R. J., and Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* 37, 361–367. doi: 10.1016/S0896-6273(02)01171-6
- Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vis.* 11, 1–34. doi: 10.1167/11.5.4
- Cary, M., and Reder, L. M. (2002). "Metacognition in strategy selection: giving consciousness too much credit" in *Metacognition: Process, function, and use*. eds. M. Izaute, P. Chambres, and P. J. Marescaux (New York, NY: Kluwer), 63–77.
- Cunningham, S. J., Turk, D. J., Macdonald, L. M., and Macrae, C. N. (2008). Yours or mine? Ownership and memory. *Conscious. Cogn.* 17, 312–318. doi: 10.1016/j.concog.2007.04.003
- De Lillo, C. (2004). Imposing structure on a Corsi-type task: evidence for hierarchical organization based on spatial proximity in serial-spatial memory. *Brain Cogn.* 55, 415–426. doi: 10.1016/j.bandc.2004.02.071
- Gao, Z., Gao, Q., Tang, N., Shui, R., and Shen, M. (2015). Organization principles in visual working memory: evidence from sequential stimulus display. *Cognition* 146, 277–288. doi: 10.1016/j.cognition.2015.10.005
- Gao, Z., Xu, X., Chen, Z., Yin, J., Shen, M., and Shui, R. (2011). Contralateral delay activity tracks object identity information in visual short-term memory. *Brain Res.* 1406, 30–42. doi: 10.1016/j.brainres.2011.06.049
- Gratton, G. (1998). The contralateral organization of visual memory: a theoretical concept and a research tool. *Psychophysiology* 35, 638–647. doi: 10.1111/1469-8986.3560638
- Guerard, K., and Tremblay, S. (2012). The effect of path length and display size on memory for spatial information. *Exp. Psychol.* 59, 147–152. doi: 10.1027/1618-3169/a000137

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Hebrew University IRB. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

HM was responsible for the design and running of the study. HM and TE performed the analysis and wrote the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02734/full#supplementary-material>

- Guerard, K., Tremblay, S., and Saint-Aubin, J. (2009). The processing of spatial information in short-term memory: insights from eye tracking the path length effect. *Acta Psychol.* 132, 136–144. doi: 10.1016/j.actpsy.2009.01.003
- Jiang, Y. V., Lee, H. J., Asaad, A., and Remington, R. (2016). Similarity effects in visual working memory. *Psychon. Bull. Rev.* 23, 476–482. doi: 10.3758/s13423-015-0905-5
- Jiang, Y., Olson, I. R., and Chun, M. M. (2000). Organization of visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 683–702. doi: 10.1037/0278-7393.26.3.683
- Kemps, E. (2001). Complexity effects in visuo-spatial working memory: implications for the role of long-term memory. *Memory* 9, 13–27. doi: 10.1080/09658210042000012
- Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006
- Luria, R., and Vogel, E. K. (2014). Come together, right now: dynamic overwriting of an object's history through common fate. *J. Cogn. Neurosci.* 26, 1819–1826. doi: 10.1162/jocn\_a\_00584
- Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356. doi: 10.1038/nn.3655
- Magen, H., and Berger-Mandelbaum, A. (2018). Encoding strategies in self-initiated visual working memory. *Mem. Cogn.* 46, 1093–1108. doi: 10.3758/s13421-018-0823-7
- Magen, H., and Emmanouil, T. A. (2018). Working memory for self-initiated and provided spatial configurations. *Q. J. Exp. Psychol.* 71, 2186–2206. doi: 10.1177/1747021817739808
- Magen, H., and Emmanouil, T. A. (2019). Estimation in self-initiated working memory for spatial locations. *Psychon. Bull. Rev.* 26, 315–324. doi: 10.3758/s13423-018-1514-x
- Milchgrub, G., and Magen, H. (2018). Self-initiated spatial working memory in young and older adults. *Memory* 26, 712–726. doi: 10.1080/09658211.2017.1402938
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/0033-295x.101.2.343

- Montoro, P. R., Villalba-Garcia, C., Luna, D., and Hinojosa, J. A. (2017). Common region wins the competition between extrinsic grouping cues: evidence from a task without explicit attention to grouping. *Psychon. Bull. Rev.* 24, 1856–1861. doi: 10.3758/s13423-017-1254-3
- Murty, V. P., DuBrow, S., and Davachi, L. (2015). The simple act of choosing influences declarative memory. *J. Neurosci.* 35, 6255–6264. doi: 10.1523/JNEUROSCI.4181-14.2015
- Orhan, A. E., and Jacobs, R. A. (2014). Toward ecologically realistic theories in visual short-term memory research. *Atten. Percept. Psychophys.* 76, 2158–2170. doi: 10.3758/s13414-014-0649-8
- Parmentier, F. B. R., Elford, G., and Maybery, M. (2005). Transitional information in spatial serial memory: path characteristics affect recall performance. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 412–427. doi: 10.1037/0278-7393.31.3.412
- Peterson, D. J., and Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychon. Bull. Rev.* 20, 1282–1289. doi: 10.3758/s13423-013-0460-x
- Peterson, D. J., Gözenman, F., Arciniega, H., and Berryhill, M. E. (2015). Contralateral delay activity tracks the influence of Gestalt grouping principles on active visual working memory representations. *Atten. Percept. Psychophys.* 77, 2270–2283. doi: 10.3758/s13414-015-0929-y
- Rossi-Arnaud, C., Pieroni, L., and Baddeley, A. D. (2006). Symmetry and binding in visuo-spatial working memory. *Neuroscience* 139, 393–400. doi: 10.1016/j.neuroscience.2005.10.048
- Schmidt, F., and Schmidt, T. (2013). Grouping principles in direct competition. *Vis. Res.* 88, 9–21. doi: 10.1016/j.visres.2013.06.002
- Slamecka, N. J., and Graf, P. (1978). The generation effect: delineation of a phenomenon. *J. Exp. Psychol. Hum. Learn. Mem.* 4, 592–604.
- Solman, G. J. F., and Kingstone, A. (2017). Arranging objects in space: measuring task-relevant organizational behaviors during goal pursuit. *Cogn. Sci.* 41, 1042–1070. doi: 10.1111/cogs.12391
- Treisman, A., and Zhang, W. (2006). Location and binding in visual working memory. *Mem. Cogn.* 34, 1704–1719. doi: 10.3758/BF03195932
- van Lamsweerde, A. E., Beck, M. R., and Johnson, J. S. (2016). Visual working memory organization is subject to top-down control. *Psychon. Bull. Rev.* 23, 1181–1189. doi: 10.3758/s13423-015-0976-3
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., et al. (2012). A century of Gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychol. Bull.* 138, 1172–1217. doi: 10.1037/a0029333
- Woodman, G. F., Vecera, S. P., and Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychon. Bull. Rev.* 10, 80–87. doi: 10.3758/BF03196470
- Xie, W., and Zhang, W. (2017). Discrete item-based and continuous configural representations in visual short-term memory. *Vis. Cogn.* 25, 21–33. doi: 10.1080/13506285.2017.1339157
- Xu, Y. D., and Chun, M. M. (2007). Visual grouping in human parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18766–18771. doi: 10.1073/pnas.0705618104

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Magen and Emmanouil. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Feature Binding of Sequentially Presented Stimuli in Visual Working Memory

Anuj Kumar Bharti<sup>1</sup>, Sandeep Kumar Yadav<sup>2</sup> and Snehlata Jaswal<sup>3\*</sup>

<sup>1</sup>Center for Biologically Inspired System Science, Indian Institute of Technology, Jodhpur, India, <sup>2</sup>Department of Electrical Engineering, Indian Institute of Technology, Jodhpur, India, <sup>3</sup>Department of Psychology, Chaudhary Charan Singh University, Meerut, India

## OPEN ACCESS

### Edited by:

Zaifeng Gao,  
Zhejiang University, China

### Reviewed by:

Shouxin Li,  
Shandong Normal University, China  
Dwight James Peterson,  
Concordia College, United States

### \*Correspondence:

Snehlata Jaswal  
sneh.jaswal@gmail.com

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 30 August 2019

**Accepted:** 07 January 2020

**Published:** 05 February 2020

### Citation:

Bharti AK, Yadav SK and Jaswal S  
(2020) Feature Binding of  
Sequentially Presented Stimuli  
in Visual Working Memory.  
*Front. Psychol.* 11:33.  
doi: 10.3389/fpsyg.2020.00033

Feature binding is a process that creates an integrated representation of an object. A change detection task with four stimuli is used to study color-shape binding of sequentially presented stimuli. Given the immense importance of locations in feature binding, and noting the confound of location information with simultaneous presentation, we compared simultaneous and sequential presentations when locations remained the same from study to test and when they changed randomly. In Experiment 1, sequential presentation implied showing the stimuli one by one to gradually build up the study display. There were no differences between the two modes of presentation in this experiment, although performance was better with unchanged locations than random locations. Experiment 2 used a sequential presentation when one stimulus vanished as the next was presented. An interaction effect showed that performance was much better with unchanged locations than random locations with simultaneous presentation, whereas locations had no effect in the sequential presentation condition. Three subsequent experiments, with drastically reduced presentation time for the display in the simultaneous presentation condition (Experiment 3), with blank intervals inserted after every stimulus in the sequential presentation condition (Experiment 4), and with a mask given immediately after the study-display presentation (Experiment 5), showed results similar to Experiment 2. Thus, we surmise that locations are a factor only in simultaneous presentation, and not in sequential presentation, and the differences between the two conditions can be attributed to post-perceptual factors within visual working memory.

**Keywords:** feature binding, simultaneous presentation, sequential presentation, locations, visual working memory

Feature binding is the process by which different characteristics, such as, orientation, size, shape, color, and location, are integrated to create an object. Binding is a necessary process for accurate perception of the world. Not only does it allow the separation of figure and ground, but also the differentiation of one object from another. Objects in the real world differ in space as well as time. Presumably, feature binding helps us differentiate objects not only when they are present together at the same time in our experience, but also when they are experienced at different times, say in a sequence. Our aim in this research is to explore the factors – whether distinct or the same – which operate in the binding of simultaneously and sequentially presented stimuli.

For testing binding in laboratory environments, a change detection task is often used. A change detection task presents a study display and a test display. Participants need to detect

whether the test display is the same or different as compared to the study display. The task can be used to test changes in uni-dimensional or multi-dimensional stimuli. When testing feature binding, all the features in the test display are the same as the study display, but their combination changes on some trials, thus the task essentially becomes “swap detection.”

Most studies of feature binding, using the swap detection task, simultaneously present the stimuli in the study display. Nevertheless, as mentioned earlier, the differentiation of objects can be over space or time. Differentiation over space (on diverse locations) is inevitable with simultaneous presentation, and separation over time yields sequential presentation. Accordingly, the array of objects in the swap detection task can be presented simultaneously or sequentially.

Simultaneous presentation of multiple objects utilizes the powerful cue of location and allows configural encoding as shown by many studies of uni-feature objects (Jiang et al., 2000; Blalock and Clegg, 2010). The importance of location in binding has been emphasized by feature integration theory (Treisman and Gelade, 1980; Treisman and Sato, 1990; Treisman, 2006; Huang et al., 2007) as well as guided search model (Wolfe, 1994). Feature integration theory (Treisman and Gelade, 1980) suggested that binding is mediated by the links of separate features to a common location. Treisman and Sato (1990) proposed that a “master map” of locations exists in our brain. Attention selects all the features associated with a particular location, and works as glue to bind those features. Neuroscientists have found the evidence for such a master map. O’Keefe and Nadel (1978) found the existence of place cells in the hippocampus. Hartley et al. (2007) supported the role of the hippocampus in topographical processing in short-term memory. Jacobs et al. (2013) did single-cell recordings from patients of epilepsy, which indicated grid cells in the entorhinal cortex and place cells in the hippocampal region. Recently, Koen et al. (2017) showed that the hippocampus plays a critical role in forming and maintaining complex bindings. Several studies have also shown that activity in the retinotopically organized sub-regions of the visual and parietal cortex is critical for visual short-term memory storage (reviewed in Xu, 2017). Behavioral studies show that location is remembered better than colors (e.g., Wheeler and Treisman, 2002). Studies also show that bindings are more vulnerable to location change and suggest that location plays a central role not only in encoding but also in maintenance and retrieval of bound objects (Treisman and Zhang, 2006; Hollingworth, 2007; Richard et al., 2008; Logie et al., 2011). Although Udale et al. (2018) provide recent evidence for strategic retrieval and decision-making by participants when task demands discourage the use of location cues, “in place” matching appears to be the default strategy of most participants even in their work. Thus, simultaneous presentation of multiple objects is considered crucial for binding by many researchers.

Nevertheless, some researchers have contrasted simultaneous and sequential modes of presentation in binding tasks. Allen et al. (2006) used a shape-color binding task with both modes of presentation. Results showed that performance was less accurate with sequential mode of presentation. Brown and Brockmole (2010) tested binding deficits in older and younger people using

simultaneous and sequential modes of presentation. Although the results did not show any effect of age on binding, performance was worse with sequential presentation for both groups.

Other research groups showed that sequential presentation is better than simultaneous presentation. Fougny and Marois (2009) used a visual working memory task in which participants had to detect changes in color, shape, either color or shape, and binding. During the retention interval, they performed a multiple object-tracking task. Results suggested that impairment caused by the secondary task was significantly reduced when objects were shown sequentially at the center of the screen. A comparison of the results of their separate experiments with simultaneous and sequential presentations shows slightly better baseline performance by participants in the sequential condition. Yamamoto and Shelton (2009) used real-life scenarios and found that sequential presentation of objects makes it easy to memorize them. They used a room layout and six different objects. Participants were shown these objects either simultaneously for 30 s or sequentially for 2.5 s per object, with the whole array being shown twice. Results showed better performance with sequential presentation. Ihssen et al. (2010) have also shown the superiority of sequential presentation. Their experiment had three conditions. In simultaneous presentation, they showed eight objects at the same time for 700 ms. In the sequential mode, they showed two displays sequentially, containing four objects at a time for 350 ms. In the third condition, the eight objects were repeated (shown twice), with each display shown for 350 ms. Results showed better memory performance in the sequential and the repeated modes.

Thus, conflicting results for different modes of presentation are observed in research studies. Simultaneous presentation increases the competition among stimuli, and the errors from within the memory set are more common than when stimuli are presented sequentially. Enrich and Ferber (2012) used eight colored squares presented either simultaneously for 400 ms, or divided into two displays of four squares each, presented sequentially. Error in detection of a particular color was more with simultaneous presentation. Haskell and Anderson (2016) also found reduced error variance with sequential rather than simultaneous presentation of circular gratings requiring judgments of orientation. Reduced errors with sequential presentation could be associated with better or equivalent performance (to simultaneous presentation), obtained at times with sequential presentation, especially in real-life conditions, where experience or familiarity with stimuli might mitigate the effects of competition and increase the distinctiveness of stimuli.

However, in experimental tasks used in the laboratories, simultaneous presentation generally yields better performance. Perhaps, this is because it allows configural encoding of the rather simple stimuli used in laboratory experiments. Stimuli can be encoded and remembered in relation to each other and form a visual pattern more easily when presented simultaneously than when presented sequentially. The relative location of stimuli is a powerful cue in simultaneous presentations. If we really wish to compare simultaneous presentation with sequential presentation, confound of location with simultaneous presentation must be removed/controlled. This is particularly important in binding studies, given the immense importance of locations in binding.

Some researchers have used a single probe at test to negate the role of location. Other researchers have attempted to control the effect of location by presenting stimuli at unchanged locations. However, because other features may be addressed through a “location map,” presenting single probes or the test stimuli at unchanged locations is not an adequate control. The huge literature on classical conditioning shows that to really break a link, it is important to randomly associate the elements participating in the link (Rescorla, 1967). To make locations irrelevant, the best strategy is to randomize them from study to test.

Thus, to unravel the effects of mode of presentation and relative locations, it seems imperative to orthogonally manipulate these two variables. In the present experiments, simultaneous and sequential presentations are compared when stimuli are presented in unchanged locations and when they are presented in random locations.

Some recent experiments studying the effect of mode of presentation on bindings with locations controlled in different ways are relevant here. Gorgoraptis et al. (2011) found that sequential presentation leads to low memory precision and more misbindings. They tested the binding of color and orientation with both modes of presentation. In the study display, they presented a number of colored bars with different orientations. In response, participants needed to adjust the orientation of the probed colored bar. The test bar was always shown at fixation. Locations were randomized in the study display in each trial in sequential as well as simultaneous presentation modes. But, in the design of this study, location was only randomized as a controlled variable, it was not an independent variable to enable an assessment of its effect in the experimental results. In another study, Pertzov and Husain (2014) using sequential presentations only compared performance in same and different locations, showing the advantage of same locations. But, in this experiment also, mode of presentation and location were not completely crossed.

Jaswal and Logie (2011) studied simultaneous and sequential modes of presentation in separate experiments keeping locations constant in one condition and randomizing locations from study to test in the other condition. Performance was inferior with sequential presentation when the participants never saw all the stimuli together in the test display, even when locations of the stimuli remained unchanged. This suggests that simultaneous presentation is better, because it gains from the relative location information concomitant with simultaneous presentation. In fact, when location was randomized, and thus rendered irrelevant to the task, there was no significant difference in performance between the simultaneous and sequential presentation experiments. Nevertheless, simultaneous and sequential presentation modes were not directly compared in their experiments and the set size at six was well beyond visual working memory capacity. Our experiments remedy this shortcoming and compare simultaneous and sequential presentations in the same experiments with set size four.

In conclusion, behavioral studies have shown equivocal results regarding performance with simultaneous and sequential presentation. In most behavioral experiments, simultaneous presentation is confounded with location information that either

encourages configural encoding (leading to better performance) or increases competition and misbinding (leading to decrement in performance). An important strategy for extricating the effects of mode of presentation and location is to manipulate both of them as separate independent variables. This is what we have done in our experiments. Five experiments are being reported here. In every experiment, simultaneous and sequential presentation modes are fully crossed with unchanged and randomized locations in a  $2 \times 2$  design. The specific aims, design, and the results of each experiment are described in the next sections.

## Participants

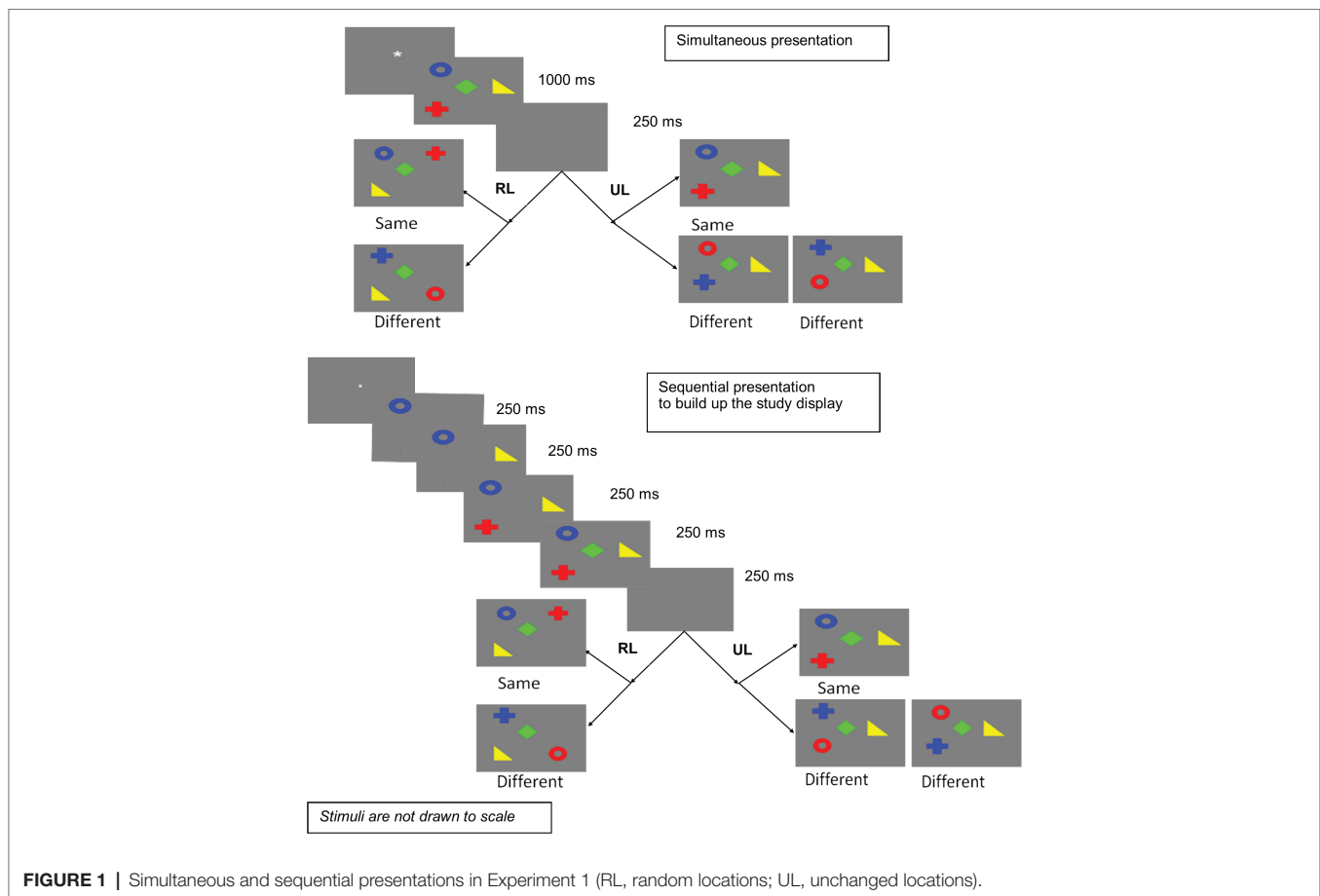
A random and independent sample of 18 participants was selected for each experiment. All experiments use a repeated measures design with both factors being within subjects. A *priori* analyses of such a design is not supported by programs such as G\*Power which estimate sample sizes. Thus, the sample size was decided on the basis of similar experiments reported in Jaswal and Logie (2011), although these experiments never compared simultaneous and sequential presentations together. They used 12 participants, so we decided to use more participants than their experiments, and recruited 18 participants in each experiment. It is pertinent here to mention that repeated measures designs are more powerful than independent samples designs. Thus, there were 90 participants in all five experiments. All participants were male undergraduates in the age range 18–22 years, reported normal or corrected to normal visual accuracy, and were paid a nominal amount as honorarium. Informed consent was taken from all participants after explaining the task, but without revealing the hypotheses.

## Apparatus and Stimuli

All experiments were designed in E Prime 2.0 (Psychology Software Tools, 2008) and were conducted on a Sony Vaio laptop with a 14 inch screen placed at a distance of about 70 cm from the participant. The screen had 100% brightness with a resolution of  $1,366 \times 768$  pixels and an Intel HD Graphics card. The four stimuli in each display were random combinations of four shapes (diamond, ring, triangle, and plus) and four colors (red, green, yellow, and blue). All stimuli were made in a square frame ( $110 \times 110$  pixels) creating a visual angle of approximately  $2.05^\circ \times 2.05^\circ$  and were presented on a gray screen in a  $3 \times 4$  invisible grid of  $338 \times 448$  pixels such that they remained in foveal vision, subtending a visual angle of approximately  $6.28^\circ \times 8.30^\circ$ .

## EXPERIMENT 1

The experiment aimed to study the effect of mode of presentation and locations on feature binding, using a change detection paradigm. We compared simultaneous and sequential presentations as the two levels of mode of presentation, and unchanged and randomized locations as the two levels of locations. As we aimed to unravel the confound between simultaneous presentation and locations, keeping these two factors as the two independent variables seemed to be a good starting point.



In this experiment, the sequential condition involves presenting the stimuli one by one, to build up the study display, as shown in **Figure 1**. Thus, in the sequential condition, an additional (temporal) cue is present. This might enhance performance in the sequential presentation condition relative to the simultaneous presentation condition. On the other hand, performance might be reduced in the sequential presentation condition relative to simultaneous presentation, if presenting stimuli one by one hampers configural encoding.

Further, unchanged locations from study to test are expected to yield better performance as compared to randomized locations, given the importance of locations as a cue in feature binding.

## Design and Procedure

The experiment was designed as a  $2 \times 2$  factorial experiment with repeated measures on both factors. The two independent variables were mode of presentation (simultaneous vs. sequential) and locations (unchanged vs. random). The trials for unchanged and random locations were mixed randomly within each block of simultaneous and sequential presentations, which were counterbalanced across participants. On half the trials comprising the unchanged locations condition, the stimuli appeared in the same locations as the study display. On the other half of the trials, comprising the random locations condition, the locations of stimuli in the test display were randomized from

the study display to the test display. **Figure 1** illustrates the design and procedure in each trial.

Each trial started with a fixation display. When the participants were ready, they pressed any key to move to the study display. The study display comprised four stimuli, which were random combinations of four colors and four shapes in each trial. The participant was to remember the bindings between colors and shapes. Simultaneous presentation implied all four stimuli presented at the same time in a single display. For sequential presentation, stimuli were presented one by one such that the display was gradually built up. Previous stimuli remained on screen as the next appeared. The study display remained on the screen for 1,000 ms for simultaneous presentation. In the sequential presentation condition, starting from the first stimulus, each next stimulus appeared after 250 ms, with all four on screen only for the last 250 ms out of a total of 1,000 ms. Thus, the total exposure duration for both presentation modes was the same. Thereafter, a blank interval was introduced for 250 ms and then a test display appeared with four stimuli. The task of the participant was to detect if any of the four stimuli changed in the binding of color and shape from the study display to the test display in each trial. The binding change happened only on 50% trials in each condition. When the change occurred, it was actually a swap between any two stimuli. Note that the



participants cannot do the swap detection task if they remembered the colors alone or shapes alone, as all the colors, and all the shapes, were repeated in the test display. Whenever a swap occurred, half the time colors changed locations, and half the time, shapes changed locations. This is experienced as different only when locations are unchanged. In the randomized locations condition, the experience of the participants does not differ for color swaps or shape swaps. The participants pressed equally separated keys for “different” and “same” to record whether they were able to detect a change in binding in each trial.

The participant had to complete the experiment in a single session. Before commencing the experiment, each participant completed 24 practice trials for each block, i.e., 48 trials in all. The experiment was completed in two blocks of 192 trials each, 384 trials in all. There was an equal number of each trial type in each block for practice as well as experimental trials. Articulatory suppression was used in each trial. The participant had to say the word “the” repeatedly from the fixation screen until after the response was given.

## Results

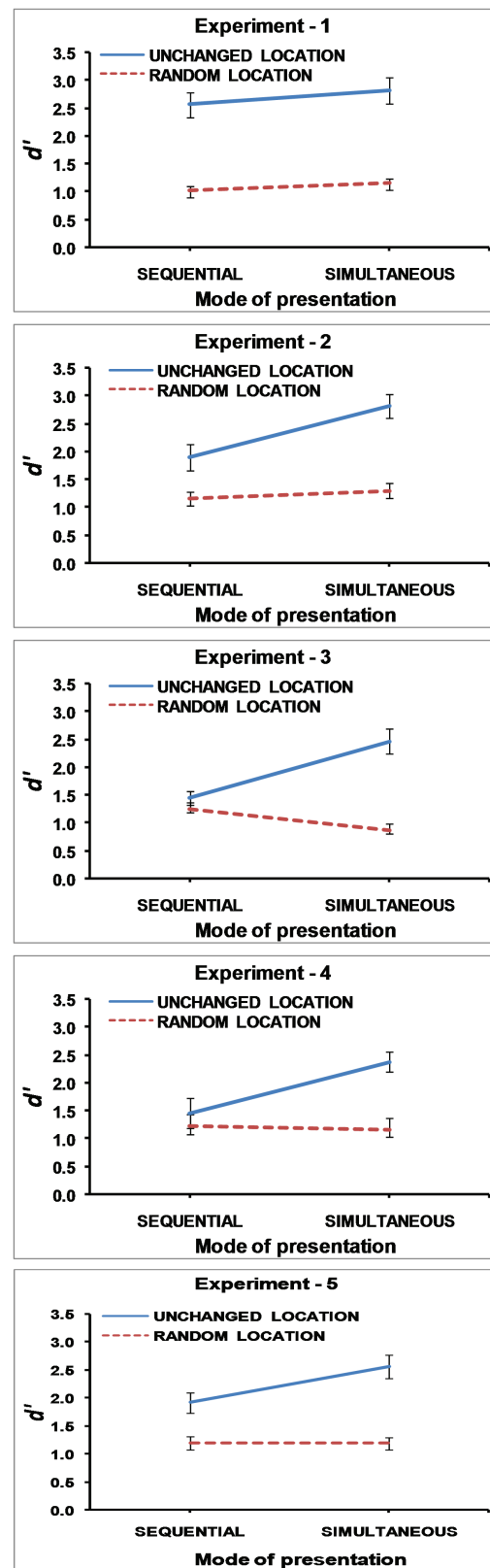
Mean change detection performance calculated from  $d'$  primes is shown in **Figure 2** for all experiments.

A repeated measures ANOVA revealed a significant main effect of unchanged and random locations,  $F(1,17) = 82.592$ ,  $MSE = 0.559$ ,  $p < 0.001$ ,  $partial \eta^2 = 0.829$ ,  $BF_{10} = 2.549 \times 10^{11}$  such that overall performance was reduced when locations were randomly changed from study to test display than when locations were unchanged. Neither the main effect of mode of presentation,  $F(1,17) = 1.089$ ,  $MSE = 0.609$ ,  $p = 0.311$ ,  $partial \eta^2 = 0.060$ ,  $BF_{01} = 3.44$ , nor the interaction effect, was significant,  $F(1,17) = 0.140$ ,  $MSE = 0.394$ ,  $p = 0.713$ ,  $partial \eta^2 = 0.008$ ,  $BF_{01} = 3.230$ . The model comprising both the main effects and the interaction effect ( $BF_{10} = 3.464 \times 10^{10}$ ) was compared with a model comprising only the main effects ( $BF_{10} = 1.119 \times 10^{11}$ ). The model comprising only the main effects better fit the data by a factor of 3.23:1.

**Table 1** shows the means of  $d'$  prime scores in all experimental conditions in this and all other experiments. **Table 2** shows the hits and **Table 3** shows the false alarms in all experiments.

## Discussion

In accordance with earlier studies, (Jaswal and Logie, 2011; Logie et al., 2011), **Figure 2** clearly shows that performance is better with unchanged locations than random locations. However, there is no significant difference between simultaneous and sequential presentation. Building up the study display by presenting stimuli one by one, and thus providing an additional temporal code, does not lead to any better performance than simultaneous presentation. This suggests that the difference between the two modes of presentation is not contingent on a temporal code alone. Perhaps other factors are more important in making simultaneous presentation better than sequential presentation. Alternatively, similar performance in the two modes of presentation may result because simultaneous presentation is also like sequential



**FIGURE 2 |** Mean  $d'$  prime scores in Experiment 1, 2, 3, 4, and 5. The error bars represent  $\pm 1$  Standard Error.

**TABLE 1 |** Mean d prime scores in all experimental conditions in the five experiments.

	Sequential presentation						Simultaneous presentation					
	Unchanged locations		Random locations		Total		Unchanged locations		Random locations		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Experiment 1	2.568	0.962	1.021	0.357	1.794	1.076	2.815	1.009	1.158	0.511	1.986	1.168
Experiment 2	1.896	1.021	1.164	0.545	1.530	0.909	2.824	0.941	1.300	0.517	2.062	1.091
Experiment 3	1.450	0.501	1.257	0.501	1.354	0.517	2.466	0.966	0.875	0.319	1.670	1.088
Experiment 4	1.456	1.161	1.218	0.932	1.336	1.075	2.382	0.749	1.153	0.543	1.768	0.910
Experiment 5	1.913	0.774	1.198	0.463	1.556	0.741	2.556	0.923	1.189	0.520	1.873	1.029

**TABLE 2 |** Hits in all experimental conditions in the five experiments.

	Sequential presentation						Simultaneous presentation					
	Unchanged locations		Random locations		Total		Unchanged locations		Random locations		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Experiment 1	0.761	0.029	0.577	0.024	0.669	0.146	0.759	0.028	0.588	0.023	0.673	0.138
Experiment 2	0.598	0.031	0.570	0.021	0.584	0.112	0.721	0.035	0.604	0.023	0.662	0.138
Experiment 3	0.598	0.023	0.614	0.029	0.606	0.111	0.706	0.033	0.600	0.027	0.653	0.137
Experiment 4	0.582	0.037	0.610	0.035	0.596	0.152	0.772	0.025	0.647	0.032	0.709	0.137
Experiment 5	0.631	0.020	0.627	0.023	0.629	0.093	0.714	0.023	0.650	0.026	0.683	0.110

**TABLE 3 |** False alarms in all experimental conditions in the five experiments.

	Sequential presentation						Simultaneous presentation					
	Unchanged locations		Random locations		Total		Unchanged locations		Random locations		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Experiment 1	0.068	0.016	0.217	0.022	0.142	0.111	0.050	0.011	0.197	0.036	0.124	0.135
Experiment 2	0.122	0.030	0.209	0.041	0.166	0.157	0.046	0.014	0.175	0.026	0.111	0.111
Experiment 3	0.153	0.033	0.199	0.034	0.176	0.143	0.077	0.024	0.288	0.031	0.182	0.157
Experiment 4	0.199	0.044	0.240	0.038	0.219	0.176	0.090	0.025	0.254	0.040	0.172	0.163
Experiment 5	0.104	0.018	0.213	0.026	0.158	0.110	0.055	0.013	0.233	0.023	0.121	0.121

presentation as participants most likely encode even simultaneously presented stimuli one by one as suggested by eye-tracking studies (e.g., Becker and Rasmussen, 2008).

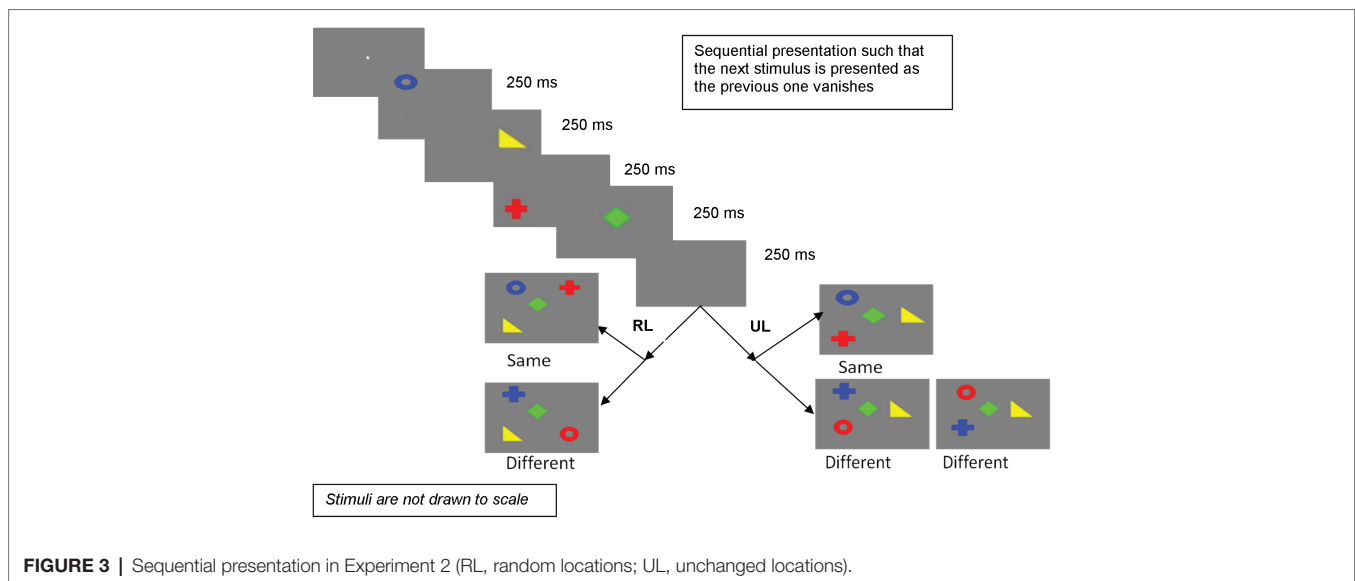
## EXPERIMENT 2

For sequential presentation in this experiment, stimuli were presented one by one such that the previous stimulus vanished as the next was presented. In such a sequential presentation, retention of the earlier stimuli becomes difficult because any given stimulus may overwrite the representation of the earlier stimuli. In the absence of previous stimuli, relational or configural encoding is much more difficult. Thus, this kind of presentation utilizes only a temporal cue in the absence of configural encoding. The performance of the participants is expected to be lesser with sequential presentation as compared to simultaneous presentation.

Further, because the representation of stimuli includes location as a feature and is thus a spatiotopic representation, feature swaps in the unchanged locations condition will be easier to detect than in the random locations condition. Also, since this spatiotopic representation is expected to exist more clearly with simultaneous presentation, therefore, the difference in performance between the unchanged and randomized locations conditions is likely to be more with simultaneous presentation rather than sequential presentation.

## Design and Procedure

The design and procedure were the same as Experiment 1, except that sequential presentation involved presenting the stimuli one by one such that the previous stimulus vanished as the next stimulus was presented. **Figure 3** depicts the procedure.



## Results

A repeated measures ANOVA revealed a significant main effect comparing unchanged and randomized locations,  $F(1,17) = 34.587$ ,  $MSE = 0.662$ ,  $p < 0.001$ ,  $\text{partial } \eta^2 = 0.670$ ,  $BF_{10} = 6.939 \times 10^5$ . Overall performance was reduced when locations were randomly changed from study to test display than when locations were unchanged. The main effect comparing simultaneous and sequential presentations was also significant,  $F(1,17) = 15.609$ ,  $MSE = 0.327$ ,  $p < 0.001$ ,  $\text{partial } \eta^2 = 0.479$ ,  $BF_{10} = 3.245$ , with performance being better with simultaneous presentation than sequential presentation of stimuli. The interaction between mode of presentation and locations,  $F(1,17) = 10.370$ ,  $MSE = 0.272$ ,  $p < 0.005$ ,  $\text{partial } \eta^2 = 0.379$ ,  $BF_{10} = 4.378$ , was also significant. **Figure 2**, which shows the mean change detection performance calculated from  $d$  primes, substantiates that the difference in performance between unchanged locations and randomized locations is much greater with simultaneous presentation,  $t(17) = 6.608$ ,  $p < 0.001$ ,  $d = 1.577$ ,  $BF_{10} = 4.607 \times 10^3$ , than with sequential presentation,  $t(17) = 3.254$ ,  $p < 0.005$ ,  $d = 0.767$ ,  $BF_{10} = 9.994$ .

To compare the results of Experiment 1 and 2, three-way analysis of variance was carried out, taking experiments as a between-subjects factor, and mode of presentation and locations as the two repeated measures factors. The main effect of experiments was not significant. However, the interaction of experiments with location,  $F(1,34) = 3.311$ ,  $MSE = 0.611$ ,  $p < 0.078$ ,  $\text{partial } \eta^2 = 0.089$ ,  $BF_{10} = 1.404$ , and the three-way interaction,  $F(1,34) = 3.136$ ,  $MSE = 0.333$ ,  $p < 0.086$ ,  $\text{partial } \eta^2 = 0.084$ ,  $BF_{01} = 1.127$ , trend toward significance. The three-way interaction was assessed by comparing the model comprising the three-way interaction and all possible main and two-way interaction effects ( $BF_{10} = 1.062 \times 10^{18}$ ) with a model comprising all three main effects and the three possible two-way interaction effects ( $BF_{10} = 1.197 \times 10^{18}$ ). The data fit better with a model without the three-way interaction only by a factor of 1.127:1. This

ratio being quite low, and the  $p < 0.084$  of the three-way interaction trending toward significance, we infer that the performance of participants is different in the two experiments.

## Discussion

The sequential presentation in this experiment presents a stimulus as the previous one vanishes. This provides a temporal cue, but does not allow configural encoding. Thus, we find that performance is not only better with simultaneous presentation, but also that within this condition, performance is better with unchanged locations, because it is in this condition that maximum advantage can be derived from configural encoding aided by the feature of locations.

Mode of presentation has a significant effect only in Experiment 2, not in Experiment 1. This implies that location is a more advantageous cue than temporal presentation for feature binding. The experiment clearly revealed the advantage of configural encoding with the aid of location information for simultaneous presentation of stimuli. The temporal cue alone is not sufficient for feature binding in the visual domain.

Since it is only in Experiment 2 that mode of presentation showed a significant difference, the further reported experiments also used sequential presentation with the previous stimulus vanishing as the next one is presented.

## EXPERIMENT 3

One of the reasons for simultaneous presentation yielding better performance than sequential presentation in Experiment 2 could be its presentation time, i.e., 1,000 ms. This presentation time was kept at 1000 ms in Experiment 2 to equate it with the total presentation time of the sequential presentation, where each of the four stimuli was presented for 250 ms. In Experiment 3, we reduced the presentation time of the study display in the simultaneous presentation condition to 250 ms to make it

equal to *one* stimulus of sequential display. Thus, one can say that participants were tested at the other logical extreme, as far as presentation time was concerned. Longer presentation time generally leads to better performance, although there are thresholds for liftoff of performance as well as when it reaches an asymptote (Busey and Loftus, 1994; Loftus and McLean, 1999). The time-based resource-sharing model of working memory (Barrouillet et al., 2004; Barrouillet and Camos, 2007) suggests that increasing the study display duration should improve performance for it allows more time for encoding and processing of stimuli. Pashler (1988) reported a significant but small increase in memory for 10 consonants presented simultaneously for 100, 300, and 500 ms. Liu and Jiang (2005) asked participants to remember objects in scene images to find that 250 ms allowed only about one object to be retained in memory. If the superior performance in simultaneous condition is indeed due to the long presentation time, reducing the presentation time of the study display in this way should drastically reduce performance in the simultaneous presentation condition, rendering it lesser than or no different from performance under the sequential presentation condition.

## Design and Procedure

The design and procedure were the same as Experiment 2 except that the study display in the simultaneous presentation condition was shown only for 250 ms.

## Results

Mean change detection scores calculated from  $d$  primes are shown in **Figure 2**. A repeated measures ANOVA revealed the main effect of unchanged and randomized locations,  $F(1,17) = 60.598$ ,  $MSE = 0.237$ ,  $p < 0.001$ ,  $partial \eta^2 = 0.781$ ,  $BF_{10} = 3.984 \times 10^4$ , in that overall performance was reduced when locations were randomly changed from study to test display than when locations were unchanged. The main effect comparing simultaneous and sequential presentations was also significant,  $F(1,17) = 7.459$ ,  $MSE = 0.242$ ,  $p < 0.014$ ,  $partial \eta^2 = 0.305$ ,  $BF_{01} = 1.226$ , with performance being better with simultaneous than sequential presentation. The interaction effect was also significant,  $F(1,17) = 23.061$ ,  $MSE = 0.381$ ,  $p < 0.001$ ,  $partial \eta^2 = 0.576$ ,  $BF_{10} = 1.760 \times 10^4$ . As depicted in **Figure 2**, there is a significant difference between unchanged and randomized locations with simultaneous presentation,  $t(17) = 7.137$ ,  $p < 0.001$ ,  $d = 1.682$ ,  $BF_{10} = 1.121 \times 10^4$ , but the difference is not significant for sequential presentation,  $t(17) = 1.406$ ,  $p = 0.178$ ,  $d = 0.331$ ,  $BF_{01} = 1.773$ .

A comparison of Experiment 2 and 3 by using a three-way ANOVA showed that neither the main effect of experiments nor any of its interactions were significant. Bayes factors were computed for each combination of main and interaction effects. A model comprising the three-way interaction with all the three main and interaction effects ( $BF_{10} = 5.608 \times 10^{16}$ ) was compared with a model of three possible main and interaction effects without the three-way interaction effect ( $BF_{10} = 7.173 \times 10^{16}$ ). The model with a three-way interaction was a slightly better fit for the data by a factor of 1.27:1.

## Discussion

The pattern of results obtained in this experiment is the same as that obtained in Experiment 2. Reducing the presentation time of the simultaneous display to a quarter of what it was in Experiment 2 had no effect on the performance of participants. Shorter exposure to the stimuli does not decrease (or increase) the performance of the participants, there being simply no significant difference between Experiment 2 and 3. These results indicate that the presentation time of the study display is not an important factor in the performance of the participants. Nevertheless, note that this experiment made changes only to the simultaneous presentation condition.

## EXPERIMENT 4

Although it seems that better performance under simultaneous presentation condition is obtained regardless of presentation time, one may argue that it is the time given for encoding the stimulus in the sequential condition, which is not enough. Ricker and Cowan (2014) tested forgetting in working memory as a function of time. They formulated the experiment comparing simultaneous and sequential conditions such that a blank interval is introduced between the stimuli in the sequential mode. Presumably, this helped in proper encoding of a stimulus, and it made performance in the sequential condition better than the performance in the simultaneous condition. Although they had tested memory for single features, analogously, we inserted blank intervals after each stimulus in the sequential presentation condition in Experiment 4, with a view to improving performance in this condition. We reasoned that blank intervals would aid consolidation or at least protect each stimulus from being overwritten by subsequent stimuli, and hence improve performance in the sequential condition.

## Design and Procedure

The design and procedure are the same as in Experiment 2 (depicted in **Figure 3**), except two related changes. In this experiment, a blank interval of 250 ms was introduced after each stimulus in the sequential presentation condition. Thus, the total time for sequential presentation becomes 1,750 ms, with four stimuli presented for 250 ms each and three blank intervals of 250 ms between the stimuli. The second change was an increase in display time for simultaneous presentation to 1,750 ms, to equalize it with the presentation time for sequential presentation. Experiment 3 (and its comparison with Experiment 2) had already shown that increasing the exposure duration has little effect on performance in the simultaneous condition. Also, a close study of Rhodes et al. (2016) showed that increasing presentation time from 900 to 2,500 ms yielded no significant difference in the retention of their participants for simultaneously presented stimuli.

## Results

Mean change detection performance calculated from  $d$  primes is shown in the **Figure 2**. A repeated measures ANOVA revealed



the main effect of unchanged and randomized locations,  $F(1,17) = 31.006$ ,  $MSE = 0.313$ ,  $p < 0.001$ ,  $\text{partial } \eta^2 = 0.646$ ,  $BF_{10} = 72.278$ , in that overall performance was reduced when locations were randomly changed from study to test display than when locations were unchanged. The main effect of simultaneous and sequential presentations is not significant,  $F(1,17) = 3.096$ ,  $MSE = 1.085$ ,  $p = 0.096$ ,  $\text{partial } \eta^2 = 0.154$ ,  $BF_{10} = 1.47$ . Nevertheless, the interaction effect was significant,  $F(1,17) = 11.826$ ,  $MSE = 0.372$ ,  $p < 0.003$ ,  $\text{partial } \eta^2 = 0.410$ ,  $BF_{10} = 6.027$ . **Figure 2** clearly depicts that the differential effect of unchanged and randomized locations is significant in the simultaneous presentation condition,  $t(17) = 8.438$ ,  $p < 0.001$ ,  $d = 1.989$ ,  $BF_{10} = 8.765 \times 10^4$ , but not in the sequential presentation condition,  $t(17) = 1.026$ ,  $p = 0.319$ ,  $d = 0.242$ ,  $BF_{01} = 2.617$ .

A comparison of Experiment 2 and 4 through a three-way ANOVA showed that neither the main effect of experiments nor any of its interactions were significant. Bayes factors were computed for all the combinations of main and interaction effects. To observe the three-way interaction effect, a model comprising the three-way interaction effect along with all the main and two-way interaction effects ( $BF_{10} = 2.587 \times 10^{10}$ ) was compared with a model of all main and two-way interaction effects only ( $BF_{10} = 7.445 \times 10^{10}$ ). The data fit better with the model without the three-way interaction effect by a factor of 2.87:1.

Another three-way ANOVA comparing Experiment 3 and 4 also did not show any differences between these experiments. Bayes factors were computed for all the possible combinations of main and interaction effects. The model comprising the three-way interaction and all the main and two-way interaction effects ( $BF_{10} = 2.897 \times 10^{10}$ ) was compared with a model comprising only the main and two-way interaction effects ( $BF_{10} = 5.256 \times 10^{10}$ ). The data fit better with the model without the three-way interaction by a factor of 1.818:1.

## Discussion

The main effect of locations and the interaction of locations and mode of presentation, both, are significant, as might be expected on the basis of the previous experiments. There is nothing new here. What is relatively more informative is that in this experiment, there is no significant difference between the two presentation modes. This might be because the overall performance in the simultaneous presentation condition decreased as compared with Experiment 2 (although the decrease does not lead to a significant main effect of experiments in the three-way ANOVA). The decrease in the performance of the participants in the simultaneous presentation condition with unchanged locations could be because the participants lost the iconic memory for the study display over the blank period. Alternatively, if the stimuli were already in the visual working memory, the participants could not sustain the relational encoding of the multiple stimuli in visual working memory. The next experiment will address whether and how far performance in this condition gains from iconic memory.

The performance of the participants in the sequential presentation condition remains the same as earlier experiments. Thus, it seems that the blank intervals, which yielded better

performance with sequential presentation of uni-feature stimuli in the experiment by Ricker and Cowan (2014), conferred no advantage in our experiment to the multi-feature sequentially presented stimuli for feature binding. Blank intervals may protect uni-feature objects from decay and interference, but have no effect on bindings.

## EXPERIMENT 5

Better performance with simultaneous presentation of stimuli may also result due to iconic memory of the visual display for simultaneous presentation, affording the correct response more easily, especially in the unchanged location condition. Iconic memory preserves the stimulus pattern for some time after it has been presented, and then visual information is transferred to visual short-term memory. Masks of different kinds have often been used to wipe out iconic memory (e.g., Sperling, 1960; Neisser, 1967; Turvey, 1973; Becker et al., 2000). Studies by Phillips (1974); Loftus et al. (1985); Loftus et al. (1992) suggest that the icon does not persist beyond the initial 100–300 ms, and in fact, longer the stimulus presentation, shorter the duration for which the icon lasts (Coltheart, 1980).

Thus, to obliterate the effects of iconic memory from performance, we decided to use a visual noise mask for 250 ms immediately after the study display in all experimental conditions, and explore whether any changes would result in the pattern of performance. Particularly, we expected that if iconic memory is the reason why simultaneously presented stimuli are better retained with unchanged locations, performance in this condition would reduce as compared to Experiment 2. However, if the stimulus representations are already in visual working memory, then they would be immune to the mask and there will be no changes in the performance of the participants, as suggested by Phillips (1974) who distinguished between sensory storage and visual short-term memory, showing that the former could be masked by noise masks, but the latter was impervious to masking. Smithson and Mollon (2006) also concluded from their study that a mask cannot penetrate higher levels of visual analysis and leaves intact the conceptual, abstract representations of stimuli.

## Design and Procedure

The design and procedure remained the same as Experiment 2. The only change was a noise mask introduced immediately after the stimulus display for 250 ms (the same duration as the study display). Thereafter, the test display was immediately presented.

## Results

Repeated measures ANOVA showed the significant main effect of the mode of presentation,  $F(1,17) = 6.949$ ,  $MSE = 0.260$ ,  $p = 0.017$ ,  $\text{partial } \eta^2 = 0.290$ ,  $BF_{01} = 1.29$ , with simultaneous presentation being better than sequential presentation. The main effect of locations was also significant,  $F(1,17) = 43.690$ ,  $MSE = 0.446$ ,  $p < 0.001$ ,  $\text{partial } \eta^2 = 0.720$ ,  $BF_{10} = 4.25 \times 10^6$ , with performance being better with unchanged locations than random locations. The interaction effect was also significant,

$F(1, 17) = 5.468$ ,  $MSE = 0.351$ ,  $p = 0.032$ ,  $partial \eta^2 = 0.243$ ,  $BF_{10} = 2.80$ . The difference of unchanged and randomized location was higher in simultaneous [ $t(17) = 5.761$ ,  $p = 0.001$ ] than sequential [ $t(17) = 3.981$ ,  $p = 0.001$ ] presentation. **Figure 2** shows the results. The similar pattern of results for Experiment 2 and 5 is clearly visible. The three-way ANOVA carried out to compare Experiment 2 and 5 showed that neither the main effect of experiments nor any of the interactions involving experiments were significant.

## Discussion

Visual noise masks were used in this experiment to eradicate the effect of iconic memory in the performance of the participants. It was of particular interest whether performance in the unchanged locations condition for simultaneously presented stimuli would reduce as compared to Experiment 2. However, there was simply no effect of the mask on the general performance level of the participants or particularly with simultaneous presentation and unchanged locations.

A three-way ANOVA was performed with Experiment 2, 3, 4, and 5 as the between-subjects factor and mode of presentation and locations as repeated measures. Neither the main effect of experiments nor any interaction of experiments with other factors was significant. Bayes factors were computed for every combination of main and interaction effects. A model comprising the three-way interaction and all the main and two-way interaction effects ( $BF_{10} = 2.026 \times 10^{27}$ ) was compared with the model with only the main and two-way interaction effects ( $BF_{10} = 7.388 \times 10^{27}$ ). The data fit better with the model without the three-way interaction by a factor of 3.64:1.

## GENERAL DISCUSSION

This research aimed to compare sequentially presented feature bindings with simultaneously presented bindings, hoping to reveal the factors, which lead to differential performance with these two modes of presentation. Previous studies, which compare simultaneous and sequential presentations, have shown mixed results, although in most studies the performance of participants is better with simultaneous presentation. We particularly designed experiments to disentangle the confound of locations with simultaneous presentation as many theories and studies have stressed the importance of locations in the process of binding (e.g., Wolfe, 1994; Treisman and Zhang, 2006; Logie et al., 2011).

The results of Experiment 1 show that merely adding a temporal cue, i.e., presenting stimuli one by one to build up the study display has no differential effect on the performance of the participants as compared to when the stimuli are simultaneously presented. Nevertheless, locations had a significant effect, with performance being significantly better when locations remained the same, than when they were randomized from study to test. This was true regardless of whether the stimuli were presented simultaneously or sequentially.

However, Experiment 2 showed a significant difference between the two modes of presentation as well as a significant interaction. In this experiment, stimuli were presented in the sequential mode of presentation such that as one stimulus was presented, the previous one vanished. Performance was worse with sequential presentation as compared to simultaneous presentation perhaps because the participants were never able to “see” the stimuli in relation to each other in the sequential presentation condition. Presumably, they were building up a mental representation of the stimuli presented in sequence, as they knew they would be tested with a whole display, having understood the experimental task, and having done many practice trials. However, in building this mental pattern/representation, it was harder for them to take advantage of the spatial relationship among the stimuli with sequential presentation such that one stimulus vanished as the next was presented. In Experiment 1, where the study display was gradually built up, they could take advantage of unchanged locations and hence the performance is not any different in the sequential presentation condition as compared with the simultaneous presentation condition.

Coming back to Experiment 2, encoding the stimuli in a configuration or pattern led to enhanced performance in the simultaneous presentation condition with unchanged locations. However, unchanged locations from study to test did not confer any advantage if the stimuli were sequentially presented. Indeed, for sequentially presented stimuli, performance was statistically not different for unchanged and random locations, indicating that location was simply not a factor in the performance of the participants with sequentially presented stimuli.

These results are in contrast to that of Pertzov and Husain (2014) who used sequential presentations of four stimuli testing the binding between color and orientation, and compared performance in same and different locations. They showed that there were less errors in the “different locations” condition. However, the differences in their experimental task as compared to ours must be noted. In their experiment, same location condition meant presenting the stimuli in exactly the same location one after the other, which does not make location unnecessary to the task, rather it makes it relevant, and thus, a factor creating confusion. In contrast, in our experiments, the stimuli are presented in different locations, which remain unchanged from study to test. Thus, in our task, locations aid in differentiating the stimuli. In their “different” locations condition, the stimuli are presented in different locations in the sequence, and tested by a probe in the center of the screen. In this case too, locations are not irrelevant to the task and the binding of other features (color and orientation in this case) may be addressed through locations, as suggested by feature integration theory (Treisman and Sato, 1990) and related studies (e.g., Treisman and Zhang, 2006; Logie et al., 2011).

Could the relatively long presentation time of 1,000 ms for all four stimuli cause better performance because in sequential presentation only 250 ms was given for each stimulus? If this was so, then giving less time to perceive stimuli in the simultaneous condition should decrease performance. However, the results of Experiment 3 showed that this was not the case.

Even reducing the presentation time of the simultaneously presented stimuli to 250 ms and thus making it equal to the presentation time of a single stimulus in the sequential condition did not affect the performance of the participants. Probably this is because all stimuli in the simultaneous presentation condition have already been encoded even at 250 ms and performance has therefore reached an asymptote. Vogel et al. (2006) have suggested that about 60 ms are required to encode the first stimulus, followed by 50 ms per stimuli for the rest of them. Although this study was with colored squares (uni-feature objects), in an earlier study, Luck and Vogel (1997) reported that the capacity of visual short-term memory is about the same for uni-feature and multi-feature objects, which is four objects. Despite suggestions that visual working memory capacity is also affected by complexity and resource demand of stimuli (Alvarez and Cavanagh, 2004; Ma et al., 2014), we believe that our four objects, which are rather simple conjunctions of color and shapes, are well within visual working memory capacity, and so presumably all stimuli in the display could be encoded within 250 ms.

Some researchers have argued that what happens in the maintenance period is as important as initial encoding; and performance is worse with sequential presentation because each stimulus gets overwritten by subsequent stimuli (Ricker and Cowan, 2014). The fourth experiment was designed to test whether introducing blank intervals after every stimulus would allow the participant to consolidate its memory and/or protect it from being overwritten by the next stimulus and hence increase the performance of the participants in the sequential presentation condition. The results did show no significant difference between sequential and simultaneous presentation conditions. However, a comparison of Experiment 2 and 4 revealed that performance did not increase in sequential presentation condition. Rather, it *decreased* in the simultaneous presentation condition with unchanged locations, probably because of the very long presentation time in this condition leading to forgetting. Thus, the blank intervals, which yielded better performance with sequential presentation in the experiment by Ricker and Cowan (2014) conferred no advantage in the sequential presentation condition in our experiment. This might be because the experiments by Ricker and Cowan were testing memory for unfamiliar shapes, whereas we were testing feature bindings. Presumably, feature bindings are already represented in the visual short-term memory beyond iconic memory, and hence do not benefit by the opportunity of consolidation (or protection) given by blank intervals to rather fragile representations of features in the initial stage of processing.

The idea that feature bindings are represented in visual short-term memory beyond iconic storage is also substantiated by Experiment 5, where we attempted to use pattern masks comprising visual noise to disrupt iconic memory representations. However, there were no significant differences in the performance of the participants as compared to Experiment 2, substantiating that feature bindings are held in the visual working memory and are thus only affected by factors, which organize information after basic perceptual processing. Supportive evidence that VSTM

representations are immune to masking is offered by several studies (e.g., Phillips, 1974; Smithson and Mollon, 2006; Sligte et al., 2008).

Consequently, we conclude that the differences between simultaneous and sequential presentations are not due to ostensible perceptual differences, but due to factors and processes that affect the organization of material/stimuli in the visual working memory. All manipulations, which could have affected perceptual processing of stimuli, viz., altering the presentation time, and inserting blank intervals after each stimulus presented in a sequence, or presenting a noise mask after the stimulus presentation, had no effect on the levels of performance of the participants. So the differential performance between simultaneous and sequential modes of presentation cannot be attributed to factors in perceptual processing. The significant interaction effect obtained in all experiments where stimuli were presented in the sequential condition such that one stimulus vanished as the next appeared substantiates that location as a feature contributes to making performance better with simultaneous presentation. The significant advantage of unchanged locations as compared to randomized locations is clear in the simultaneous presentation condition in all experiments. It is clear that this advantage accrues only when stimuli can be encoded in relation to each other, being presented together in multiple locations.

However, in the case of stimuli presented sequentially, location is simply not relevant to performance as keeping it the same or randomizing it has no effect on the performance of the participants. Perhaps this is because these stimuli are already represented in visual working memory. This idea is further substantiated by the last three experiments, which show that the performance in the sequential presentation condition is immune to manipulations designed to alter the encoding of stimuli such as changes to presentation time, or inserting blank intervals, or using a noise mask immediately after stimulus presentation. Also, as suggested by one of the reviewers, performance in the sequential presentation condition could have been worse because participants were required to maintain items for a longer duration in this condition, particularly in the experiment where blank intervals were inserted. Clearly, this difficulty in “maintenance” would occur only if the stimuli were already present in visual working memory. In sum, we speculate that sequences are encoded or consolidated into visual working memory relatively automatically and perhaps sooner as compared to simultaneously presented stimuli. Analogous to the advantage that sentences have over lists in verbal working memory due to long-term knowledge (Allen et al., 2018), perhaps sequences of visual stimuli too benefit from temporal cues which are simply absent in simultaneously presented stimuli. Alternatively, competition among simultaneously presented stimuli may act as a bottleneck and retard the progress of these early visual representations into working memory. This idea is supported by the experimental finding that differences between simultaneous and sequential presentations are evident only at larger set sizes and are not shown with set sizes within working memory capacity (Igel and Harvey, 1991; Dent and Smyth, 2006). Another explanation could be that participants are using different strategies to process simultaneously and sequentially presented stimuli.

Udale et al. (2018) have argued that participants can use different strategies to encode and process stimuli when required by task demands in the absence of locations being relevant. In fact, they also suggest individual differences among participants in the use of these strategies. Much further research is required to explore exactly which factors and processes in visual working memory are relevant for binding sequentially presented stimuli.

On the basis of current studies, it may be concluded that while performance with simultaneous presentation relies on location information, performance with sequential presentation is relatively immune to presence/absence of location information. It is also clear that post-perceptual processes within visual working memory are presumably responsible for the differences in performance due to simultaneous and sequential presentation.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## REFERENCES

- Allen, R. J., Baddeley, A., and Hitch, G. J. (2006). Is the binding of visual features in working memory resource demanding? *J. Exp. Psychol. Gen.* 135, 298–313. doi: 10.1037/0096-3445.135.2.298
- Allen, R. J., Hitch, G. J., and Baddeley, A. D. (2018). Exploring the sentence advantage in working memory: insights from serial recall and recognition. *Q. J. Exp. Psychol.* 71, 2571–2585. doi: 10.1177/1747021817746929
- Alvarez, G. A., and Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol. Sci.* 15, 106–111. doi: 10.1111/j.0963-7214.2004.01502006.x
- Barrouillet, P., Bernardin, S., and Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *J. Exp. Psychol. Gen.* 133, 83–100. doi: 10.1037/0096-3445.133.1.83
- Barrouillet, P., and Camos, V. (2007). "The time-based resource-sharing model of working memory" in *The cognitive neuroscience of working memory*. eds. N. Osaka, R. H. Logie, and M. D'Esposito (New York: Oxford University Press), 59–80.
- Becker, M. W., Pashler, H., and Anstis, S. M. (2000). The role of iconic memory in change-detection tasks. *Perception* 29, 273–286. doi: 10.1068/p3035
- Becker, M. W., and Rasmussen, I. P. (2008). Guidance of attention to objects and location by long term memory of natural scenes. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1325–1338. doi: 10.1037/a0013650
- Blalock, L. D., and Clegg, B. A. (2010). Encoding and representation of simultaneous and sequential arrays in visuospatial working memory. *Q. J. Exp. Psychol.* 63, 856–862. doi: 10.1080/17470211003690680
- Brown, L., and Brockmole, J. (2010). The role of attention in binding visual features in working memory: evidence from cognitive ageing. *Q. J. Exp. Psychol.* 63, 2067–2079. doi: 10.1080/17470211003721675
- Busey, T. A., and Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychol. Rev.* 101, 446–469. doi: 10.1037/0033-295X.101.3.446
- Coltheart, M. (1980). The persistencies of vision. *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* 290, 57–69. doi: 10.1098/rstb.1980.0082
- Dent, K., and Smyth, M. M. (2006). Capacity limitations and representational shifts in spatial short term memory. *Vis. Cogn.* 13, 529–572. doi: 10.1080/13506280444000760
- Emrich, S., and Ferber, S. (2012). Competition increase binding errors in visual working memory. *J. Vis.* 12, 1–16. doi: 10.1167/12.4.12
- Fougnie, D., and Marois, R. (2009). Attentive tracking disrupts feature binding in visual working memory. *Vis. Cogn.* 17, 48–66. doi: 10.1080/13506280802281337

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee, Department of Psychology, Chaudhary Charan Singh University, Meerut. The participants provided their written informed consent to participate in this study.

## AUTHOR'S NOTE

AB carried out this research as part of PhD. He was supported for PhD by a fellowship from the Ministry of Human Resource Development, India.

## AUTHOR CONTRIBUTIONS

AB, SY, and SJ together conceptualized the study and wrote the paper.

- Gorgoraptis, N., Catalao, R., Bays, P., and Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *J. Neurosci.* 31, 8502–8511. doi: 10.1523/JNEUROSCI.0208-11.2011
- Hartley, T., Bird, C. M., Chan, D., Cipolotti, L., Husain, M., Vargha-Khadem, F., et al. (2007). The hippocampus is required for short-term topographical memory in humans. *Hippocampus* 17, 34–48. doi: 10.1002/hipo.20240
- Haskell, C., and Anderson, B. (2016). Attentional effects on orientation judgements are dependent on memory consolidation processes. *Q. J. Exp. Psychol.* 69, 2147–2165. doi: 10.1080/17470218.2015.1105830
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 31–47. doi: 10.1037/0096-1523.33.1.31
- Huang, L., Treisman, A., and Pashler, H. (2007). Characterizing the limits of human visual awareness. *Science* 317, 823–825. doi: 10.1126/science.1143515
- Igel, A., and Harvey, L. O. (1991). Spatial distortions in visual perception. *Gestalt Theory* 13, 210–231.
- Ihssen, N., Linden, D. E. J., and Shapiro, K. (2010). Improving visual short-term memory by sequencing the stimulus array. *Psychon. Bull. Rev.* 17, 680–686. doi: 10.3758/PBR.17.5.680
- Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X. X., et al. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. *Nat. Neurosci.* 16, 1188–1190. doi: 10.1038/nn.3466
- Jaswal, S., and Logie, R. H. (2011). Configural coding in visual feature binding. *J. Cogn. Psychol.* 23, 586–603. doi: 10.1080/20445911.2011.570256
- Jiang, Y., Olson, I. R., and Chun, M. M. (2000). Organization of visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 683–702. doi: 10.1037/0278-7393.26.3.683
- Koen, J. D., Borders, A. A., Petzold, M. T., and Yonelinas, A. P. (2017). Visual short-term memory for high resolution associations is impaired in patients with medial temporal lobe damage. *Hippocampus* 27, 184–193. doi: 10.1002/hipo.22682
- Liu, K., and Jiang, Y. (2005). Visual working memory for briefly presented scenes. *J. Vis.* 5, 650–658. doi: 10.1167/5.7.5
- Loftus, G. R., Duncan, J., and Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 530–549.
- Loftus, G. R., Johnson, C. A., and Shimamura, A. P. (1985). How much is an icon worth? *J. Exp. Psychol. Hum. Percept. Perform.* 11, 1–13. doi: 10.1037//0096-1523.11.1.1
- Loftus, G. R., and McLean, J. E. (1999). A front end to a theory of picture recognition. *Psychon. Bull. Rev.* 6, 394–411. doi: 10.3758/BF03210828



- Logie, R. H., Brockmole, J. R., and Jaswal, S. (2011). Feature binding in visual short term memory is unaffected by task irrelevant changes of location, shape, and colour. *Mem. Cogn.* 39, 29–36. doi: 10.3758/s13421-010-0001-z
- Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846
- Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356. doi: 10.1038/nn.3655
- Neisser, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- O' Keefe, J., and Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- Pashler, H. (1988). Familiarity and visual change detection. *Percept. Psychophys.* 44, 369–378. doi: 10.3758/BF03210419
- Pertsov, Y., and Husain, M. (2014). The privileged role of location in visual working memory. *Atten. Percept. Psychophys.* 76, 1914–1924. doi: 10.3758/s13414-013-0541-y
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Atten. Percept. Psychophys.* 16, 283–290. doi: 10.3758/BF03203943
- Psychology Software Tools (2008). *E-prime 2.0 [computer software]*. Pittsburgh, PA: Psychological Software Tools.
- Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychol. Rev.* 74, 71–80. doi: 10.1037/h0024109
- Rhodes, S., Parra, M. A., and Logie, R. H. (2016). Ageing and feature binding in visual working memory: the role of presentation time. *Q. J. Exp. Psychol.* 69, 654–668. doi: 10.1080/17470218.2015.1038571
- Richard, A. M., Luck, S. J., and Hollingworth, A. (2008). Establishing object correspondence across eye movements: flexible use of spatiotemporal and surface feature information. *Cognition* 109, 66–88. doi: 10.1016/j.cognition.2008.07.004
- Ricker, T. J., and Cowan, N. (2014). Differences between presentation methods in working memory procedures: a matter of working memory consolidation. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 417–428. doi: 10.1037/a0034301
- Sligte, I. G., Scholte, H. S., and Lamme, V. A. (2008). Are there multiple visual short-term memory stores? *PLoS One* 3:e1699. doi: 10.1371/journal.pone.0001699
- Smithson, H., and Mollon, J. (2006). Do masks terminate the icon? *Q. J. Exp. Psychol.* 59, 150–160. doi: 10.1080/17470210500269345
- Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr. Gen. Appl.* 74, 1–29. doi: 10.1037/h0093759
- Treisman, A. (2006). How the deployment of attention determines what we see. *Vis. Cogn.* 14, 411–443. doi: 10.1080/13506280500195250
- Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5
- Treisman, A., and Sato, S. (1990). Conjunction search revisited. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 459–478. doi: 10.1037//0096-1523.16.3.459
- Treisman, A., and Zhang, W. (2006). Location and binding in visual working memory. *Mem. Cogn.* 34, 1704–1719. doi: 10.3758/BF03195932
- Turvey, M. T. (1973). On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli. *Psychol. Rev.* 80, 1–52. doi: 10.1037/h0033872
- Udale, R., Farrell, S., and Kent, C. (2018). Task demands determine comparison strategy in whole probe change detection. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 778–796. doi: 10.1037/xhp0000490
- Vogel, E. K., Woodman, G. F., and Luck, S. J. (2006). The time course of consolidation in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1436–1451. doi: 10.1037/0096-1523.32.6.1436
- Wheeler, M. E., and Treisman, A. M. (2002). Binding in short-term visual memory. *J. Exp. Psychol. Gen.* 131, 48–69. doi: 10.1037/0096-3445.131.1.48
- Wolfe, J. M. (1994). Guided search 2.0- a revised model of visual search. *Psychon. Bull. Rev.* 1, 202–238. doi: 10.3758/BF03200774
- Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends Cogn. Sci.* 21, 794–815. doi: 10.1016/j.tics.2017.06.013
- Yamamoto, N., and Shelton, A. (2009). Sequential versus simultaneous viewing of an environment: effects of focal attention to individual objects' location on visual spatial learning. *Vis. Cogn.* 17, 457–483. doi: 10.1080/13506280701653644

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bharti, Yadav and Jaswal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Metacognitive Perspective of Visual Working Memory With Rich Complex Objects

Tomer Sahar<sup>1,2\*</sup>, Yael Sidi<sup>1</sup> and Tal Makovski<sup>1</sup>

<sup>1</sup> Department of Psychology and Education, The Open University of Israel, Ra'anana, Israel, <sup>2</sup> Department of Psychology, University of Haifa, Haifa, Israel

## OPEN ACCESS

### Edited by:

Hagit Magen,  
The Hebrew University of Jerusalem,  
Israel

### Reviewed by:

Marian Berryhill,  
University of Nevada, Reno,  
United States  
Mark Nieuwenstein,  
University of Groningen, Netherlands

### \*Correspondence:

Tomer Sahar  
tomelico@gmail.com

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 10 October 2019

**Accepted:** 27 January 2020

**Published:** 25 February 2020

### Citation:

Sahar T, Sidi Y and Makovski T  
(2020) A Metacognitive Perspective  
of Visual Working Memory With Rich  
Complex Objects.  
Front. Psychol. 11:179.  
doi: 10.3389/fpsyg.2020.00179

Visual working memory (VWM) has been extensively studied in the context of memory capacity. However, less research has been devoted to the metacognitive processes involved in VWM. Most metacognitive studies of VWM studies tested simple, impoverished stimuli, whereas outside of the laboratory setting, we typically interact with meaningful, complex objects. Thus, the present study aimed to explore the extent to which people are able to monitor VWM of real-world objects that are more ecologically valid and further afford less inter-trial interference. Specifically, in three experiments, participants viewed a set of either four or six memory items, consisting of images of unique real-world objects that were not repeated throughout the experiment. Following the memory array, participants were asked to indicate where the probe item appeared (Experiment 1) whether it appeared at all (Experiment 2) or whether it appeared and what was its temporal order (Experiment 3). VWM monitoring was assessed by subjective confidence judgments regarding participants' objective performance. Similar to common metacognitive findings in other domains, we found that subjective judgments overestimated performance and underestimated errors, even for real-world, complex items held in VWM. These biases seem not to be task-specific as they were found in temporal, spatial, and identity VWM tasks. Yet, the results further showed that meaningful, real-world objects were better remembered than distorted items, and this memory advantage also translated to metacognitive measures.

**Keywords:** subjective judgment, real-world objects, confidence, meaning, appearance errors

## INTRODUCTION

To what degree does one have access to own mental processes? The body of research termed *Metacognition* aims to answer this question. The field of metacognition refers to "thinking about thinking" (Flavell, 1979) and it deals with the evaluation and monitoring of cognitive processes and the control and regulation of these processes (see Koriat, 2007 for review). Broadly speaking, monitoring of cognitive processes refers to one's awareness of the operation of a specific cognitive process while it occurs. From an experimental perspective, monitoring is usually assessed by collecting participants' direct (subjective) confidence judgments regarding the relevant process, and matching that with the actual outcome. Consequently, metacognitive studies often find that monitoring may be based on heuristics: a pragmatic but not necessarily optimal approach to

generate subjective judgments, as in certain situations, they may prove unreliable and lead to biased decisions (Tversky and Kahneman, 1974).

To assess the degree to which monitoring coincides with the actual performance, two central measures are used: calibration and resolution (Fleming and Lau, 2014; Fiedler et al., 2019). Calibration, or absolute accuracy, refers to the gap between subjective confidence judgments and task performance scores (e.g., correct responses). Thus, calibration is maximized when the proportion of correct responses equals to the subjective confidence judgments given by the observer, and the absolute difference is zero. That is, subjective confidence ratings equal to the actual performance. An overconfidence bias occurs when subjective confidence exceeds task scores—as the observer overestimates her performance. Conversely, an underconfidence bias occurs when high performance is underestimated.

Resolution, or relative accuracy, is the extent to which confidence judgments vary between a correct or incorrect response. This is measured as a correlation between confidence and accuracy. Resolution is maximized when high performance is predicted by high confidence judgments and low performance is predicted by low confidence judgments (Ackerman and Goldsmith, 2011, for reviews, see Schwartz and Eklides, 2012; Goldsmith et al., 2014). Note that calibration and resolution are two independent measures. Calibration reflects the extent of deviation from being subjectively accurate in confidence judgments, whereas resolution is a correlation that reflects the extent of how judgments represent and change with performance.

The current study aimed at examining visual working memory (VWM) from a metacognitive perspective. VWM is considered to be a fundamental, capacity-limited on-line buffer, and individual differences in this ability are related to high cognitive functions, such as intelligence (Luck and Vogel, 2013). Hence, understanding how people access and assess the content held in VWM can shed new light on the mechanisms underlying VWM processes. Furthermore, the relationship between working memory and metacognitive abilities is likely to be bi-directional. For example, Komori (2016) showed that in a dual-task setting, observers with high working memory capacity made more accurate judgments about their performance than observers with low capacity. On the flip side, researchers are also relying on the assumption that observers have accurate metacognitive reports and use that to assess VWM processes (Adam et al., 2017). Thus, studying metacognitive processes within VWM can gain valuable insights into both VWM and metacognitive processes.

Metacognitive studies of VWM have mainly examined the accuracy of subjective estimations of VWM limit and the extent that subjective and objective visual knowledge dissociate from one another. The correspondence of objective VWM measures and subjective judgments showed that, overall, subjective judgments reliably reflect (at least to some extent) VWM content and objective visual information (Rademaker et al., 2012; Vandenbroucke et al., 2014; Samaha and Postle, 2017; Suchow et al., 2017). Yet, other studies have stressed the separability of objective visual information and subjective judgments (Bona et al., 2013; Bona and Silvanto, 2014; Vlassova et al., 2014; Maniscalco and Lau, 2015). For instance, Adam and Vogel (2017)

showed that while subjective judgments predicted some variation in memory performance, observers were consistently unaware of their own memory failures.

One issue of measuring metacognitive processes in VWM is the repeated use of a limited set of simple stimuli (e.g., colors, orientations) in VWM tasks. This results in a narrow, homogeneous stimuli space and increases the likelihood of proactive interference. The outcome of proactive interference is that items from previous trials are harder to reject, and are mistakenly reported as if they appeared in the current trial (e.g., Keppel and Underwood, 1962; Hartshorne, 2008; Makovski and Jiang, 2008; Makovski, 2016; but see Lin and Luck, 2012). Thus, without accounting for these errors, studies might inaccurately estimate VWM performance, and more importantly for the current purposes, they might impair our ability to adequately assess the metacognitive processes involved in VWM because both subjective and objective performance are likely to be contaminated by information encountered in previous trials.

One way to minimize proactive interference is by using real-world objects instead of simple stimuli. These stimuli afford to test numerous distinct items without repetition throughout the experiment (Endress and Potter, 2014; Makovski, 2016; Shoval et al., 2019). Testing real-world objects in VWM tasks further bears an ecological benefit as we typically interact with meaningful, rich, complex objects and not with impoverished stimuli such as color patches. Accordingly, recent findings showed that the visual and semantic heterogeneity of meaningful objects leads to an improved VWM performance and extend the typical limit of VWM capacity (Brady et al., 2016; Shoval et al., 2019). However, it is still unknown how accurate people are in monitoring VWM of rich, real-world objects.

The goal of the current study was to explore observers' ability to monitor VWM processes using distinct complex stimuli and various VWM tasks. Three experiments were conducted in order to reveal the correspondence between objective and subjective memory performance while minimizing proactive interference by using non-repeating images of real-world objects. Specifically, we measured observers' resolution and calibration while they were performing VWM tasks with unique (i.e., presented only once in the task) and distinct real-world objects. This allowed us to estimate the metacognitive abilities of VWM across three domains (e.g., spatial, identity, temporal) with minimal interference from the information shown in previous trials.

## EXPERIMENT 1

The aim of the first experiment was to examine spatial VWM performance from a metacognitive perspective. Thus, on each trial, participants memorized a set of six images of real-world objects, presented sequentially at distinct locations (Makovski, 2016). After a short retention period, one of the presented images appeared and participants were asked to indicate the item's location. Next, they were asked to evaluate their confidence by indicating the degree of certainty that they chose the correct item's location on a 0–100 scale. This allowed us to assess

both subjective and objective performance and thereby estimate resolution and calibration.

## Method

### Participants

Participants were students (age: 18–35) from the Open University of Israel who took part in the experiment for course credit. All had normal or corrected-to-normal visual acuity and were without learning disabilities or attention disorders. Power calculation showed that a minimum sample size of 20 participants provided a power of 0.8 for detecting a Cohen's  $d$  effect size of 0.66 using a two-tailed paired samples  $t$ -test. Twenty-two participants completed Experiment 1 (19 females, mean age = 27).

### Materials and Stimuli

The task was created and implemented with MATLAB software (MathWorks Inc., Natick, MA, United States, 2010) and Psychtoolbox (Brainard, 1997) on a 23.5" Eizo Foris monitor (1920 × 1080, 120 Hz) and a standard PC. Participants were tested individually in a dim room. They sat approximately 50 cm from the screen. A black fixation cross ( $0.96^\circ$ ) was presented at the center of a white background screen. Two columns of three black-frame empty squares ( $5.6^\circ \times 5.6^\circ$ ) served as place-holders (located  $14^\circ$  to the left and right of fixation, and  $14^\circ$  above, at fixation level, and  $14^\circ$  below the fixation, **Figure 1**). The image set included 1200 images of real-world objects ( $4.8^\circ \times 4.8^\circ$ ) drawn from a previously published set (Brady et al., 2008<sup>1</sup>). Confidence judgments were collected by scrolling with the mouse over a rectangle bar ( $40^\circ \times 1.9^\circ$ ). The initial position of the cursor was at the middle of the bar (i.e., at 50%). The bar was interactively filled with the color blue from its left edge to the position of the cursor. The percentage of the filled area, from 0 to 100, served as a numeric indicator for confidence and it was presented above the rectangle. Participants finalized their judgment response by pressing the space key. Note that responding without moving the cursor was impossible, and a response of 50% was not allowed.

### Procedure

The trial began with a 950 ms fixation and place-holders display that remained visible throughout the trial. Each trial consisted of six unique images, randomly drawn in each trial for each subject. Each image appeared in isolation within a distinct place-holder for 500 ms. The items appeared sequentially in random order and after the last image was shown, a fixation cross was displayed for 600 ms. Then, the probe item, which was always one of the six items presented in that trial, appeared above fixation together with the six empty place-holders and the mouse cursor at fixation. The probe item was evenly and randomly chosen from the six possible locations and six serial positions. Participants were instructed to indicate the place-holder in which the probe item appeared by clicking on its position using the mouse. There was no time limit for this task and only accuracy was emphasized. After a response was registered, participants were instructed to indicate their subjective confidence that they made a correct response by scrolling with the mouse over a

0 ("not-sure") to 100 ("very sure") scale. A numeric value of confidence was accordingly shown, and the participants were instructed to choose any value that reflected their subjective confidence except for 50%. The next trial began after 500 ms of a blank display (**Figure 1A**).

Participants performed 180 experimental trials (five trials in each of the six locations, six serial-positions combinations), preceded by eight practice trials. Every 36 trials participants could take a short break.

## Results

### Accuracy

**Figure 1B** depicts performance as a function of serial position. The overall accuracy reflected moderately poor performance, but was above chance level [16.6%,  $M = 44.3\%$ ,  $SD = 22.9$ ,  $t(21) = 18.2$ ,  $p < 0.001$ ]. A repeated-measures analysis of variance (ANOVA, Greenhouse–Geisser corrected) of accuracy as a function of the probed-item serial position was significant,  $F(2.84, 59.7) = 82.063$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.796$ . Bonferroni corrected comparisons showed that accuracy was best for the last item to (the 6th item, all  $p$ 's  $< 0.001$ ). The second to last item (the 5th item) was also better than all previous positions (all  $p$ 's  $< 0.001$ ). There was no other significant difference between positions 1–4 (all  $p$ 's  $> 0.1$ ) except that the fourth item was better than the second item ( $p = 0.017$ ). These results reflect a typical recency effect as the locations of the last two items were better remembered than the location of the first four items (Broadbent and Broadbent, 1981).

### Confidence

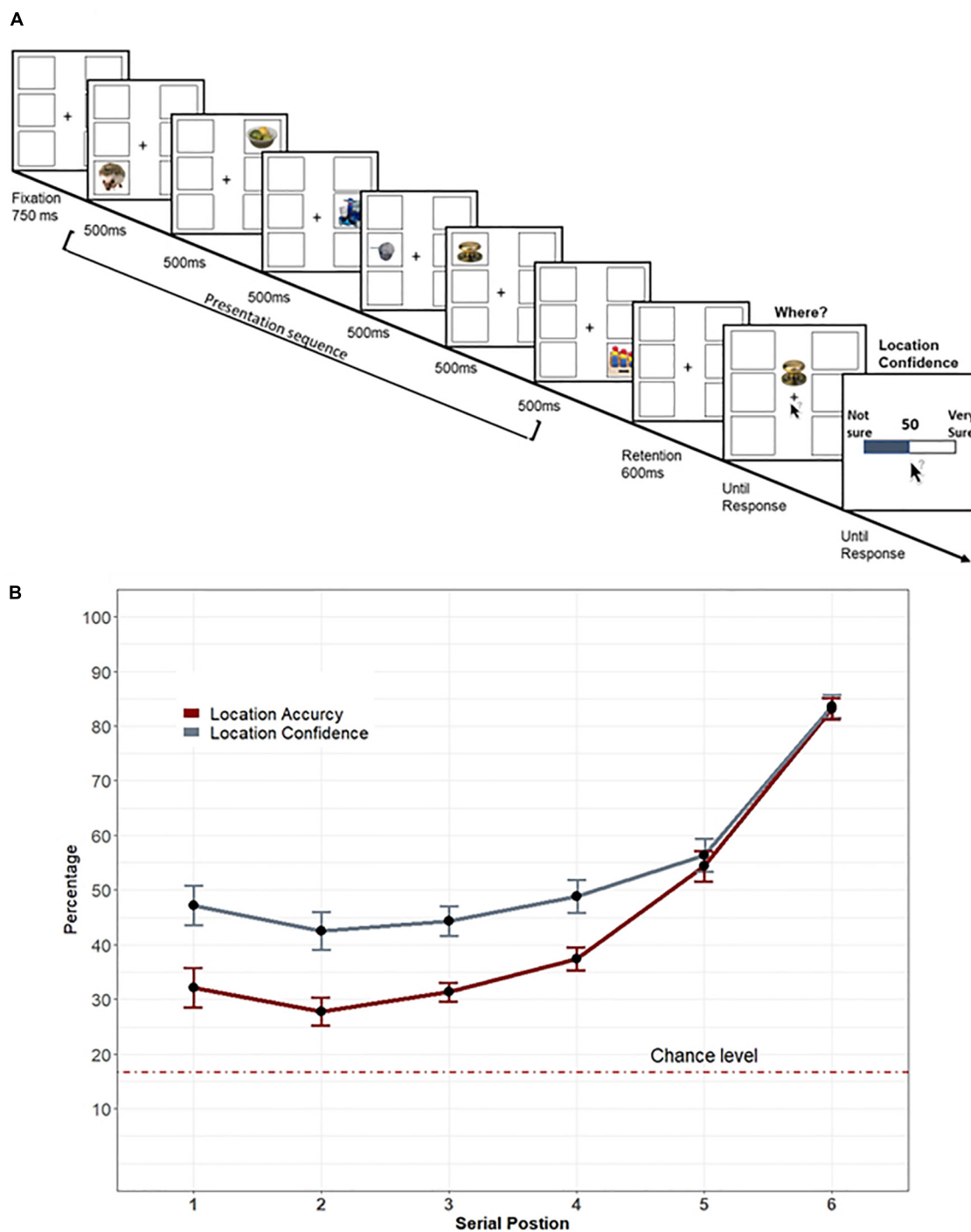
Similar to accuracy, a repeated-measures ANOVA of confidence ratings (**Figure 1B**) as a function of the probed-item serial position revealed a significant effect,  $F(2.79, 58.75) = 77.872$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.788$ . Bonferroni corrected *post hoc* comparisons showed that confidence was largest for the last presented item (all  $p$ 's  $< 0.001$ ). The confidence of the second to last presented item was also larger than all previously presented items (all  $p$ 's  $< 0.046$ ). No other significant difference was found (all  $p$ 's  $> 0.08$ ).

### Calibration

Calibration was calculated as the difference between confidence and accuracy in each serial position of each subject. Repeated-measures ANOVA of calibration as a function of serial position revealed a significant main effect,  $F(5, 105) = 9.063$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.301$ . To further examine the source of the overconfidence bias, *post hoc* Bonferroni corrected comparisons showed that the last item significantly differed from the third, second, and first items (all  $p$ 's  $< 0.033$ ). The fifth item differed from all previous items (all  $p$ 's  $< 0.028$ ). No other comparisons were significant. Bayesian one-sample  $t$ -test further showed a reliable and positive difference from zero for the first four items (BF = 36, 13, 21, 5.4, respectively), but did not show a reliable difference from zero for the last two presented items (BF = 0.26, 0.22, respectively). This suggests that the overconfidence bias was driven from the first four items, whereas observers were well-calibrated for the last two items (see **Figure 1B**).

<sup>1</sup> <https://bradylab.ucsd.edu/stimuli.html>





**FIGURE 1 | (A)** Illustration of Experiment 1's trial's sequence: Each trial began with blank place-holders. The presentation sequence consisted of six items, each at a distinct location. After the final stimulus and between the presentation of the probe, a blank place-holders display was shown for 600 ms. Following this short retention, participants were asked to indicate where the probed item appeared and to rate their confidence regarding their response. **(B)** Experiment 1's results: Mean confidence (gray line) and the mean percentage of correct location (red line) plotted as a function of the probed-item's serial position during the presentation sequence. Error bars represent standard error of the mean.

## Resolution

For each participant, a resolution was calculated as the Gamma correlation coefficient (i.e., Goodman–Kruskal correlation)

between accuracy and confidence (Nelson, 1984) collapsed across all serial positions. The averaged resolution across participants was moderate ( $M = 0.521$ ,  $SD = 0.1$ ), suggesting that observers'

discrimination between the better- and less-remembered location of the probed item was only accurate to some extent.

## Discussion

The results of Experiment 1 showed that participants' sensitivity to their performance in the VWM task was moderate—as reflected by their resolution. However, this estimation (0.521) seems to be numerically larger than correlations previously reported in other metacognitive studies of VWM, which varied between 0.19 and 0.47 (0.22–0.39, Thomas et al., 2012; 0.19–0.43, Yue et al., 2013; 0.47, Adam and Vogel, 2017; but see Masson and Rotello, 2009). We also found that the calibration was highly influenced by the item's serial position as observers were overconfident in the first four items but well-calibrated in the last two items.

In the current experiment, we asked observers about the item's location and not about the memory of the item itself. That is, the objective and subjective measures were only based on the spatial memory of the item (where the item was presented). However, a crucial aspect of memory is the explicit access to the item's identity, which is also often used as a measure of memory performance (e.g., “was this chair presented?”). Therefore, in Experiment 2, we turn to directly examine whether participants explicitly remember the probed item and particularly their confidence that the probed item appeared.

## EXPERIMENT 2

While we usually interact with both the item's identity and its location, they are not necessarily recalled together nor do they decay together in an obligatory manner (Köhler et al., 2001; Pertzov et al., 2012). Thus, testing spatial memory alone, as in the previous experiment, does not provide a full view of the metacognitive abilities of VWM. Specifically, it remains unclear whether people can accurately assess their VWM when it is based on the item's identity.

Several changes were therefore done in Experiment 2. First, the presentation set-size was reduced to four items to ensure that the capacity limit was not exceeded. As in Experiment 1, each item appeared at a distinct location, and items were not repeated throughout the experiment. After the presentation sequence, observers were asked to indicate whether they explicitly remember that the probed item appeared and to rate their confidence regarding the item's appearance. Importantly, the probed item was always an item from the presentation sequence. Afterward, they were asked to indicate its location. When participants reported that the probed item did not appear, an “appearance error” was registered but the trial continued the same. That is, participants were asked to guess a possible location and were not told anything about whether the item actually appeared or not (note that the item always appeared). This allowed us examine the location accuracy in those trials where participants reported that they do not remember that the item appeared (i.e., its identity). Note that in this experiment, we focused on participants' reports and confidence ratings that the item appeared and thus we did not measure the

confidence in knowing where the item appeared (as was in Experiment 1).

## Method

### Participants

Thirty-six new participants from the Open University of Israel completed Experiment 2 (25 females, mean age 25.3).

### Materials, Stimuli, and Procedure

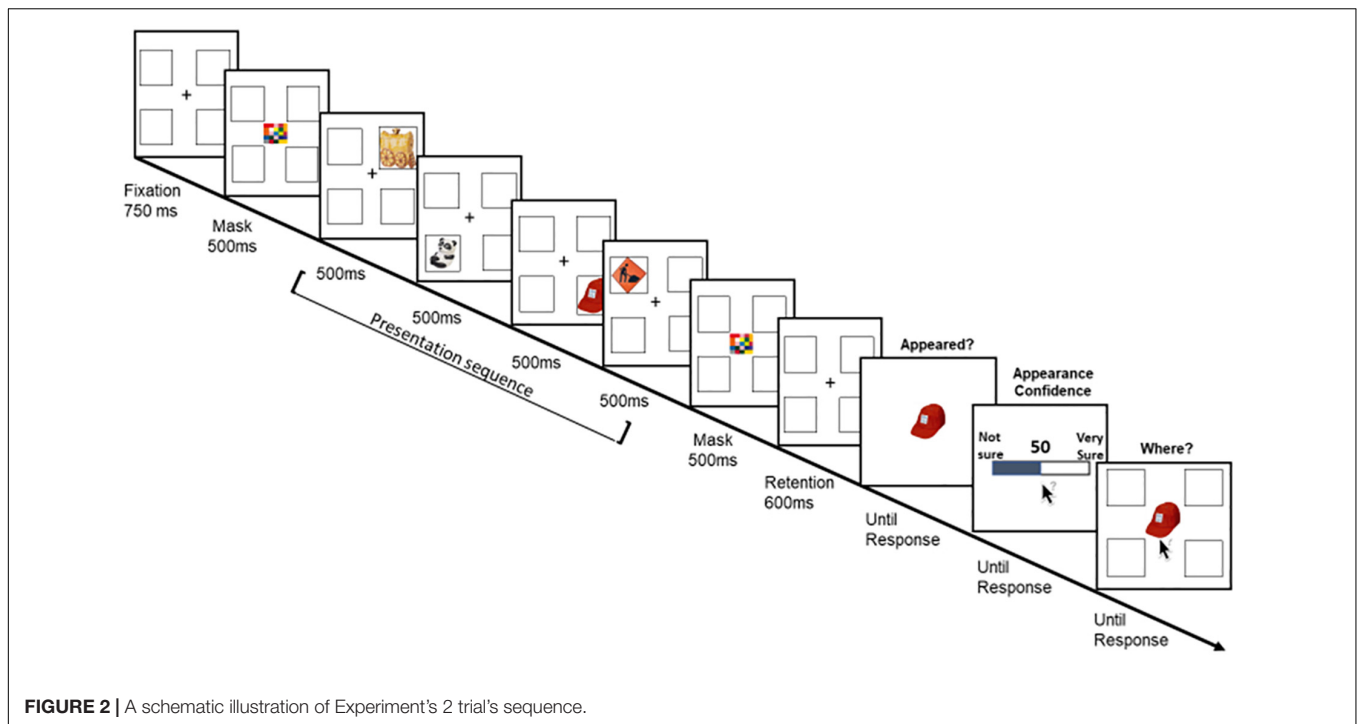
Unique non-repeated images were drawn from the same set used in Experiment 1. Because memory set-size was reduced, and to avoid verbal coding, articulatory suppression was included in the task. Specifically, each trial began with a randomly drawn word (out of 24 Hebrew words, three letters, two syllables) that participants were asked to repeat aloud throughout the presentation sequence. Participants initiated the trial by pressing the mouse button when they were ready. Then, the fixation with the place-holders display was shown for 750 ms. The place-holders display consisted of four black-frame squares (2 by 2,  $5.6 \times 5.6$  each)  $12^\circ$  to the left and right of fixation and  $12^\circ$  above and below fixation. To match the presentation condition to Experiment 1 (in terms of forward and backward masking), a multi-colored square ( $4.8 \times 4.8$ ) was presented for 500 ms at the center of the display before the first item and after at the fourth item. After 600 ms of a retention interval, the probed item appeared, and participants were asked to indicate whether or not they remember this item by pressing the keyboard keys “1” or “2,” respectively. Importantly, the probed item was always one item from the sequence. Note that when participants reported that the item did not appear, the trial continued the same. After this response, participants were asked to rate their subjective confidence that the probed item appeared or not using the same method in Experiment 1. After their appearance confidence rating was registered, the probed item re-appeared at the center of the display together with the empty place-holders and the participants were asked to use the mouse to indicate the probed-item location. If participants indicated that the probed item did not appear, they were asked to guess its possible location. No confidence question was asked for this response (Figure 2).

Participants performed eight practice trials before continuing to 192 experimental trials (12 trials in each of the four locations, four serial positions combinations). In order to control for items' memorability, all participants viewed the same presentation sequences and probed items, in the same order. Every 32 trials participants could take a short break and the session lasted approximately 30 min. In all other respects, the method was identical to Experiment 1.

## Results

### Statistical Analyses

Due to the subjective nature of “appearance” (i.e., remembering an item), different participants produced different proportions (if any) of “not appeared” errors (as on each trial the probe item was always presented), range = 0–42%, median = 9.9%, SD = 10.04. To avoid excluding participants on the basis of balancing sample-sizes, we used mixed-effects models to analyze the data of “not-appeared” trials. For each outcome variable (i.e., appearance



confidence, appearance calibration, location accuracy), we used a simple linear mixed-effects model (LMM) with only one main effect (serial position: 1–4) as a fixed effect and participants as a random effect.

The effects in this model were tested using the *lme* function of the *nlme* package (version 3.1 – 137, Pinheiro et al., 2019). The *F*-values and *p*-values (approximation by the degrees of freedom) of the effects were calculated by implementing the *ANOVA* function from the *stats* package (version 3.5.2, R Core Team, 2019). *Post hoc* comparisons are reported with Tukey adjustments.

For “appeared” trials (without the problem of uneven observations), we used, as before, repeated-measures ANOVAs with serial position as a within-subject factor and Greenhouse–Geisser corrected where needed. *Post hoc* comparisons were adjusted using Bonferroni correction. **Figure 3** depicts the results.

### Proportion of Trials and Appearance Errors

Overall, in about 89% of trials, participants reported correctly that they remember the probed item (**Figure 3**, left). The complement ( $M = 11\%$ ) reflects the proportion of appearance errors (**Figure 3**, right) which can be viewed as “misses” because the probed item was actually presented on each trial. These errors varied with serial position,  $F(3,105) = 14.26$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.29$ . Bonferroni corrected *post hoc* comparisons showed that the last item was the most accurate (reported as appeared, all  $p$ 's  $< 0.008$ ). The third item differed from the second item ( $p = 0.012$ ). No other comparisons were significant. We now turn to describe the memory and metacognitive measures as a function of participants' reports on the item's appearance.

### Reported Appeared

#### Appearance confidence

Appearance confidence was high ( $M = 92.4$ ,  $SD = 9.1$ ) and although numerically similar, differences in confidence as a function of the probed-item serial position were significant,  $F(1.9,69.4) = 18.9$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.351$ . *Post hoc* comparisons showed that the confidence for the last item was the highest (all  $p$ 's = 0.001). The third item was higher than the second item ( $p = 0.004$ ). The first item was also higher than the second ( $p = 0.041$ ).

#### Appearance calibration

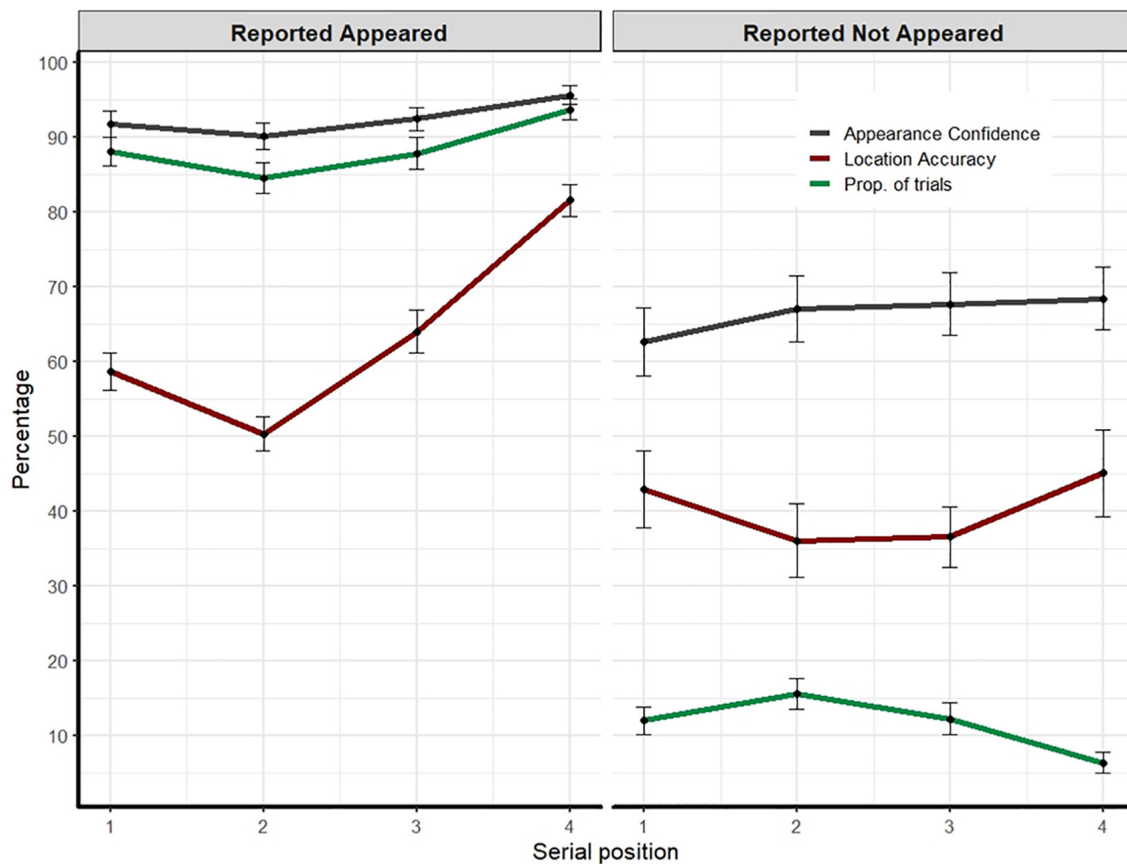
There was no significant effect of the item's serial position on calibration  $F(3,105) = 2.005$ ,  $p = 0.118$ . Overall, one-sample two-sided Bayesian *t*-test showed the average calibration was numerically close to zero but was inconclusively different than zero [ $M = 3.9$ ,  $SD = 13.7$ , two-sided one-sample *t*-test,  $t(35) = 1.7$ ,  $p = 0.095$ ,  $BF = 0.675$ ]. Nevertheless, the lack of overall overconfidence bias in these data should be taken with caution because of a ceiling effect as the performance was quite high.

#### Location accuracy

Overall, the averaged location memory was moderate ( $M = 63.6\%$ ,  $SD = 10.5$ ) and a similar pattern of serial position emerged  $F(2.4,85.5) = 46.8$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.573$ . The last item was the most accurate (all  $p$ 's  $< 0.001$ ). The third item was better than the second item ( $p < 0.001$ ). The first item was better than the second ( $p = 0.002$ ).

### Reported Not Appeared

We now turn to examine the appearance confidence and calibration, as well as the location accuracy, separately for the



**FIGURE 3 |** Experiment's 2 results: Appearance confidence, mean correct location responses, and the proportion of trials as a function of participants' appearance report (i.e., appear, did not appear) and the item's serial position. Error bars represent one standard error of the mean.

trials on which participants erroneously reported that the probed item did not appear.

#### Appearance confidence

The overall confidence that the probed item did not appear was relatively high ( $M = 66.2$ ,  $SD = 26.01$ ) and did not differ across serial positions,  $F(1,87) < 1$ ,  $p > 0.1$ .

#### Appearance calibration

There was no significant effect of the item's serial position  $F(1,87) = 1.43$ ,  $p = 0.234$ . Importantly, overall, a one-sample two-sided Bayesian  $t$ -test showed the average difference of appearance confidence from the proportion of errors was positively different than zero ( $M = 52.8$ ,  $SD = 24.6$ , two-sided Wilcoxon test,  $V = 7627$ ,  $p < 0.001$ ,  $BF > 1e^{+8}$ ), suggesting that participants exhibited high degree of confidence that the probed item did not appear.

#### Location accuracy

Overall, location accuracy was poor ( $M = 40.2\%$ ,  $SD = 29.6$ ), but above chance level (25%, one-sided Wilcoxon test,  $V = 4951$ ,  $p < 0.001$ ). There was no significant effect of serial position,  $F(1,87) < 1$ ,  $p > 0.1$ .

#### Appearance Resolution

We calculated the Gamma correlation coefficient for appearance responses and appearance confidence, across all trials (i.e., the correct response was every trial that was reported as "appeared"). One participant's data were excluded due to ceiling performance as none of the trials was reported as "not-appeared." Overall, the resolution was high ( $M = 0.69$ ,  $SD = 0.25$ ) but should be taken cautiously due to the high proportion of correct responses and ceiling performance.

#### Discussion

In this experiment, we asked participants whether they remember that the probed-item appeared. When they did remember the item, they were fairly calibrated (and slightly overconfident). In addition, both identity and location memory showed serial position effects, as the last item was the most accurate in all respects.

More importantly, we found that in about 11% of all trials, participants erroneously reported with relatively high confidence that the probe item did not appear. This number might not seem high; however, these appearance errors (or more likely rapid-forgetting errors) were found even though the conditions were optimal for remembering that the probed item appeared. Namely,



memory load was low and within capacity, the items were presented in isolation at distinct locations for a relatively long duration, and there were no intrusions from previous trials. Thus, these results imply that observers can easily fail to remember a visually distinct and fully visible item even though VWM capacity is not exceeded. Furthermore, the relatively high confidence of subjects that the probed item did not appear, although admittedly falls below the confidence in “appeared” trials, demonstrates the underestimation of memory errors. That is, rather than being less confident that the item appeared, participants were more confident that the item did not appear.

These identity-appearance errors further point to the fragility of VWM. However, intriguingly, the location accuracy of those trials was above chance level, suggesting that participants had at least some degree of access to the item’s memory representation or that location memory was more accessible. This notion is consistent with the finding that in change-detection paradigms, observers are quite good at localizing the change but show difficulties in trying to report the identity of the changed item (Caplovitz et al., 2008; Hughes et al., 2012). Yet, before making strong conclusions in that regard, several alternative explanations should be considered. First, it may be possible that in some trials participants mistakenly reported that they did not remember the probed item, when in fact they did and were thus able to report its correct location. Second, participants’ strategy might account for some of these results. For example, participants might choose the location based on an “educated guess” by elimination if they remember more than one item and its location, or alternatively, they can choose an “empty” location—one that is not associated with any remembered item.

It is also possible that an old-new type question (i.e., item appeared or not) is more difficult and requires access to more information than the question of where the item appeared. There are more possibilities to choose from when trying to judge whether an item appeared (comparing the item against all possible memory traces) than making a decision regarding its location (one out of four possible locations, but see Makovski et al., 2010)<sup>2</sup>. Thus, participants might be prone to report “did not appear,” as it is harder to access the item’s identity, but its representation (and location information) is still accessible to some extent.

## EXPERIMENT 3

The results thus far showed that participants overestimate their VWM abilities in knowing where an item was presented (Experiment 1) and in knowing that an item was, in fact, presented (Experiment 2). In the final experiment, we wished to replicate these findings and extend them to another dimension of the task: the temporal domain.

Thus, the design of this experiment included subjective and objective questions about both the identity (appearance) and the temporal order of the probe. In order to prevent observers from using spatial locations as memory cues, each item was

presented (separately) at the center of the screen. After the four items were presented, participants were now asked to report, first, whether they remembered the probed item and to rate their confidence on appearance, and second, what was the item’s serial position (i.e., was the probed item shown first, second, third, or fourth? Note that the item was always shown in the presentation sequence) and again to indicate their confidence in their temporal order response (**Figure 4B**). Same as in Experiment 2, the trial continued irrespective of the appearance response. Thus, we were able to also measure the temporal order subjective and objective measures in trials where participants did not remember the item’s identity.

The second goal of Experiment 3 was to test the role of semantics in the metacognitive processes of VWM. Indeed, a notable difference between complex, real-world objects and simple stimuli is that semantic meaning might be involved in VWM tasks, particularly when using real-world objects (Shoval et al., 2019). The role of meaning might be especially important in metacognition as reporting about explicit subjective judgments, while the probe is presented, could be biased by item’s label and meaning (e.g., “I haven’t seen this car”). Therefore, in order to directly test whether meaning plays a role in VWM and metacognition of VWM, two types of items were tested in Experiment 3: images of intact objects (high-semantic) and distorted versions of the same images (low-semantic).

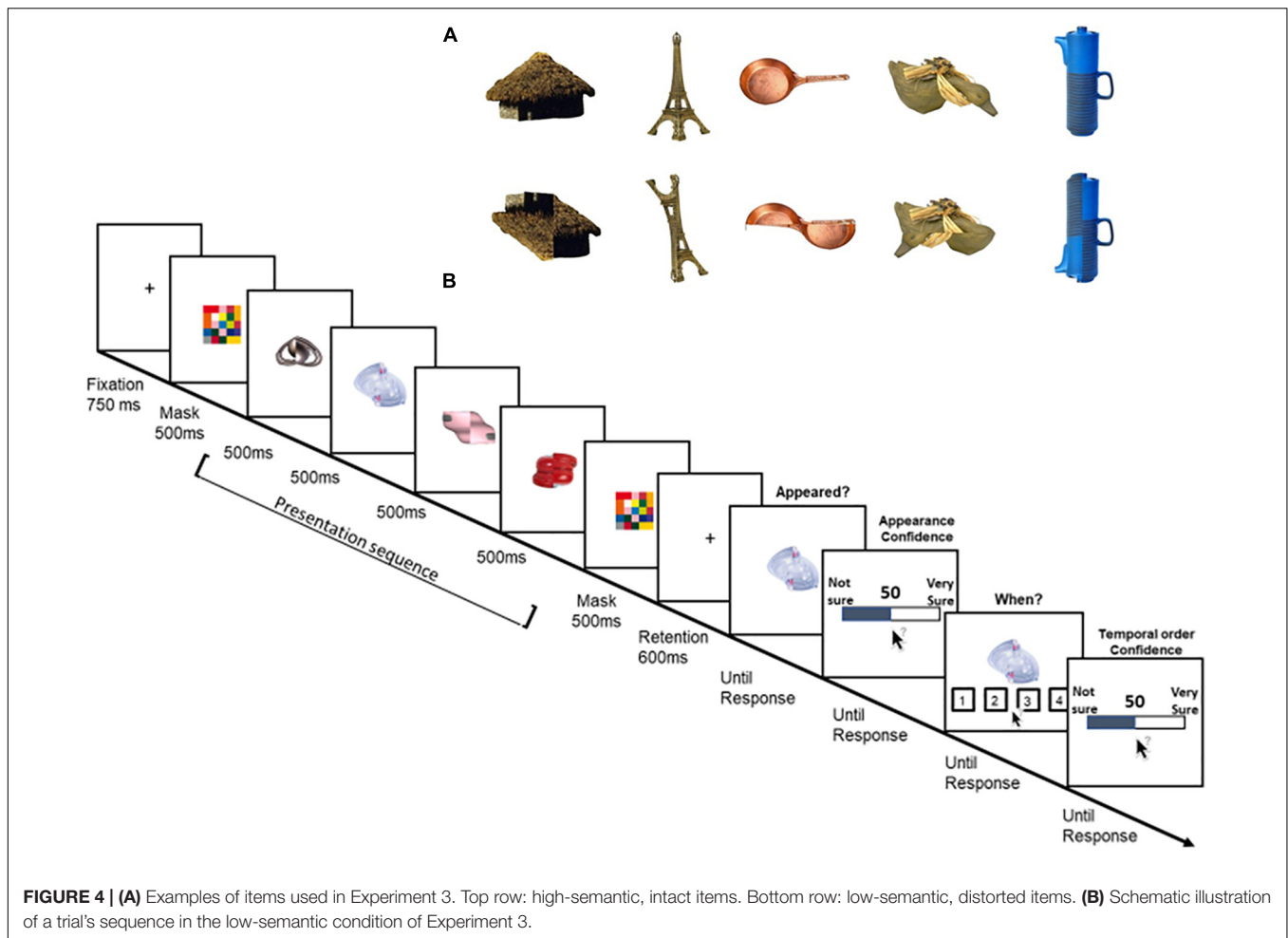
Specifically, images of real-world objects were flipped 90° along their vertical or horizontal midline (**Figure 4A**). This simple manipulation kept most of the item’s visual properties but reduced its meaning. Indeed, there might not be fully meaningless object, and one might still “recognize” the identity of the distorted item (e.g., **Figure 4A**, the Eiffel tower) or attribute a meaning to what seems to be meaningless (e.g., the distorted pan, **Figure 4A**). However, in a previous study, several manipulation checks showed that participants were slower to verbally name these distorted objects, and these items were rated as less “meaningful” than their intact counterparts (Makovski, 2018). Furthermore, these low-semantic items were shown to considerably reduce VWM capacity (Shoval and Makovski, submitted) and were, therefore, good candidates to test the role of semantic in the metacognition of VWM.

## Method

The high-semantic objects were 600 images drawn from the same set of Experiment 1 (**Figure 4A**, top). The low-semantic objects were distorted versions of those images (**Figure 4A**, bottom). Specifically, half of the image was flipped along the vertical or horizontal midline. This allowed to disrupt the item’s meaning but to keep the visual statistics similar for both intact and distorted items (for further details and manipulation checks, see Makovski, 2018. The full stimuli set is publicly available at <https://osf.io/3rn9k/>).

Each trial began with the presentation of a black fixation cross against a white background for 750 ms. Then, a multicolored square was shown for 500 ms at the center of the screen followed by the memory items. The memory sequence included four items that were shown sequentially at the center of the screen, each for 500 ms. At the end of the sequence, the multicolored square was

<sup>2</sup>We thank Dr. Dominique Lamy for raising this possibility.



shown again for 500 ms before a 600 ms blank retention interval. Then, the probed item appeared at the center of the screen and participants were asked to indicate whether they remember that this item appeared, by pressing the keys “1” if they did or “2” if not. Note that same as in Experiment 2, the trial continued regardless the appeared or not response. Next, participants were asked to rate their confidence regarding that item’s appearance, in the same method as in Experiment 1. After this response was registered, the probed item re-appeared together with four black numbered frames (each  $4.5^\circ \times 4.5^\circ$ , located below fixation). The numbers (1,2,3,4 from left to right) were shown inside the frames and represented the serial position. Participants were asked to indicate the item’s serial position using the mouse. Then, they were asked to rate their confidence that they were correct in a similar way as before (**Figure 4B**).

Participants performed two 96-trials blocks (i.e., each serial position was tested 24 times). Each block consisted of either the high-semantic items or the low-semantic items. The starting order of the blocks was counterbalanced across participants. Before the task started, participants performed eight practice trials, four in each semantic condition. Every 32 trials, participants could take a short break. Twenty-nine new participants completed Experiment 2 (23 females, mean age 25.5).

## Results

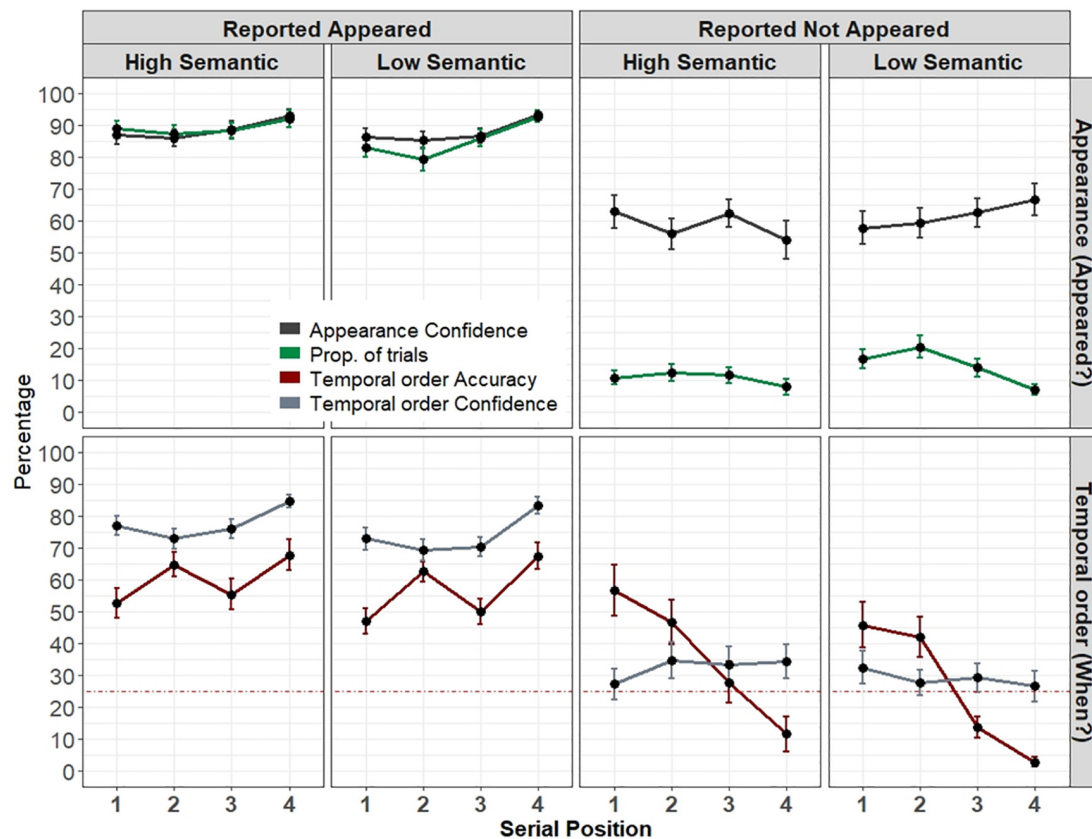
### Statistical Analyses

As before, for “appeared” trials, we used repeated-measures ANOVA with semantic level (high, low) and serial position (1–4) as a within-subject factor with Greenhouse–Geisser correction where needed. *Post hoc* comparisons were Bonferroni corrected. For the “not appeared” trials, we used LMM approach, similar to Experiment 2. The model for Experiment 3 was with two main effects: semantic (high, low) and serial position (1–4) and an interaction term as fixed effects, and participants as a random effect.

**Figure 5** depicts the results.

### Proportion of Trials and Appearance Errors

Overall, the proportion of “appeared” trials was high ( $M = 87.2\%$ ,  $SD = 12.1$ ). The complement, the proportion of appearance errors (12%) was similar to Experiment 2. Note that same as in Experiment 2, all probes were actually always presented on each trial. A repeated-measures ANOVA revealed a main effect of semantic-level  $F(1,28) = 6.7$ ,  $p = 0.015$ ,  $\eta_p^2 = 0.19$ , in that high-semantic items were more often reported as “appeared” ( $M = 89.1\%$ ,  $SD = 13.4$ ) than low-semantic items ( $M = 85.3\%$ ,  $SD = 14.7$ ). There was also a significant main



**FIGURE 5 |** Experiment's 3 results. Top: Proportion of trials and appearance confidence as a function of semantic level, reported appeared or not, and serial position. Bottom: Percentage of correct temporal order responses and mean confidence rating as a function of the item's semantic level, reported appeared or not, and serial position. The dotted line represents the 25% chance level. Error bars represent one standard error of the mean.

effect of serial position,  $F(2.07, 58.17) = 11.57$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.29$ . *Post hoc* comparisons showed that the last position was the most accurate with the highest proportion of trials (all  $p$ 's  $< 0.007$ ). These two factors interacted,  $F(3, 84) = 5.04$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.15$ . The interaction was mostly driven by a smaller proportion of "appeared" responses (and a larger proportion of appearance errors) for low semantic items that appeared in the second position ( $p = 0.008$ ). Low-semantic items also produced a more pronounced serial effect, with fewer appearance errors for the last item than all other items (all  $p$ 's  $< 0.02$ ) and fewer errors in the third item than the second ( $p = 0.035$ ). These effects were not present in the high-semantic items as no other comparison reached significance.

### Reported Appeared

#### Appearance confidence

The confidence that the item appeared was high ( $M = 88$ ,  $SD = 12$ ). The same repeated-measures ANOVA with semantic level (high, low) and serial position (1–4) as factors, showed only a serial position effect  $F(1.85, 51.84) = 17.52$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.38$ . *Post hoc* comparisons showed that only the last item's appearance confidence was higher than all other positions (all  $p$ 's  $< 0.001$ ).

#### Appearance calibration

Participants were well-calibrated ( $M = 1.07$ ,  $SD = 9$ ) and the calibration was not statistically different from zero [two-sided one-sample  $t$ -test,  $t(28) = 0.637$ ,  $p = 0.529$ ,  $BF = 0.238$ ]. The analysis showed a main effect of semantic-level,  $F(1, 28) = 4.69$ ,  $p = 0.039$ ,  $\eta_p^2 = 0.14$ , as participants were better calibrated for high-semantic items. There was also a significant interaction with serial position,  $F(3, 84) = 4.07$ ,  $p = 0.013$ ,  $\eta_p^2 = 0.12$ , as participants were slightly overconfident in the first and second items of the low-semantic items ( $M = 3.2$ ,  $M = 6$ ), but were slightly underconfident in those items for high-semantic items ( $M = -2.1$ ,  $M = -1.2$ ). It should be noted again, however, that any conclusion about confidence bias in these results should be taken cautiously given the high-performance in the identity memory task.

#### Temporal order accuracy

Overall, the percentage of correct temporal order responses was moderate and above chance level [ $M = 58.4\%$ ,  $SD = 16.5$ ,  $t(28) = 10.89$ ,  $p < 0.001$ ]. There was only a significant effect of serial position,  $F(3, 84) = 10.22$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.26$ . *Post hoc* comparisons showed that the last item was more accurate than

any of the previous items (all  $p$ 's < 0.001) except for the second item. The second item was more accurate than the third and first items ( $p$ 's < 0.029). Further analysis suggested that this advantage of the second position possibly stems from a response bias (i.e., a large proportion of "second position" responses,  $M = 35\%$ ,  $SD = 9.2$ , likely because participants mostly guessed "2" whenever they were unsure).

### Temporal order confidence

The overall temporal order confidence was relatively high ( $M = 75.7$ ,  $SD = 14$ ). The analysis showed a main effect of semantic-level,  $F(1,28) = 7.5$ ,  $p = 0.011$ ,  $\eta_p^2 = 0.21$ . The confidence was higher for high-semantic items ( $M = 77.5$ ) than low-semantic items ( $M = 73.8$ ). The analysis also showed a significant serial position effect,  $F(2,159.7) = 25.29$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.14$ , in that the last item received the highest confidence (all  $p$ 's < 0.001).

### Temporal order calibration

Similar to the spatial domain tested in Experiment 1, we found an overconfidence bias in the temporal domain as calibration was positively above zero [ $M = 17.3$ ,  $SD = 12.5$ ,  $t(28) = 7.4$ ,  $p < 0.001$ ,  $BF > 1e^{+5}$ ]. The analysis further showed a main effect of serial position,  $F(2,3,65.2) = 9.68$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.25$ , resulting from the accurate calibration of the second position.

### Reported Not Appeared

We now turn to examine the appearance confidence and calibration, as well as the temporal order accuracy, confidence, and calibration separately for the trials on which participants erroneously reported that the probed item did not appear.

### Appearance confidence

Overall, the confidence that an item did not appear was relatively high ( $M = 66.2$ ,  $SD = 26$ ). There were no significant effects of neither semantic-level and serial position, and the two did not interact, all  $F$ 's < 1.3, all  $p$ 's > 0.1.

### Appearance calibration

The overall difference between the confidence that the probed item did not appear, and the proportion of these errors was positively high and different than zero ( $M = 44.7$ ,  $SD = 25.9$ ,  $V = 400$ ,  $p < 0.001$ ,  $BF > 1e^{+5}$ ). That is, participants exhibited a high degree of confidence that the probed item did not appear. Again, there were no significant differences in both semantic and serial position, nor significant interaction, all  $F$ 's < 2.1, all  $p$ 's > 0.1.

### Temporal order accuracy

The overall temporal order accuracy was poor ( $M = 32.7\%$ ,  $SD = 37.35$ ) and was not different than chance level (25%, Wilcoxon one-sided test,  $V = 7185.5$ ,  $p = 0.054$ ). The analysis showed a significant effect of serial position  $F(3,127) = 14.3$ ,  $p < 0.001$ . *Post hoc* comparisons showed that the last item was worse than all other items (all  $p$ 's < 0.001) except for the third item. The third item was worse than the second and the first items ( $p < 0.006$ ). No other comparison was significant.

### Temporal order confidence

The confidence was overall low ( $M = 30.9$ ,  $SD = 26.5$ ). The analysis showed that none of the factors were significant, all  $F$ 's < 1, all  $p$ 's > 0.1.

### Temporal order calibration

The overall calibration ( $M = -1.8$ ,  $SD = 44$ ) was not statistically different than zero (two-sided Wilcoxon test,  $V = 5318$ ,  $p = 0.9$ ,  $BF = 0.234$ ). Yet, this does not suggest that participants were well-calibrated because there was also a robust serial position effect  $F(3,127) = 12.122$ ,  $p < 0.001$ . *Post hoc* comparisons showed that last and third items ( $M = 24$ ,  $M = 10$ ) largely differed from the second and first items ( $M = -13$ ,  $M = -20$ , all  $p$ 's < 0.016). Thus, the change from overconfidence to underconfidence (which evens out to zero) was driven only by the large decrease in the temporal order accuracy across serial positions (probably because of the tendency to guess 1 or 2 when not knowing) that was not accompanied by a change in confidence (see Figure 5, bottom right).

### Appearance resolution

We calculated the Gamma correlation coefficient for appearance responses and appearance confidence separately for each semantic condition. Four participants' data were excluded due to ceiling performance, as only one or no trials were reported as "not-appeared." Overall and similar to Experiment 2, the resolution was high ( $M = 0.69$ ,  $SD = 0.27$ ) and should be taken cautiously due to ceiling performance. There was also no difference between high-semantic ( $M = 0.70$ ,  $SD = 0.28$ ) and low-semantic items [ $M = 0.68$ ,  $SD = 0.28$ , paired two-sided  $t$ -test,  $t(24) = 0.88$ ,  $p = 0.384$ ].

### Temporal order resolution

We also calculated the Gamma correlation coefficient for temporal order responses and its confidence across all trials, separately for each semantic condition. Overall, the temporal order resolution was high and similar to Experiment 1 ( $M = 0.48$ ,  $SD = 0.22$ ). There was also a significant difference between high-semantic items ( $M = 0.51$ ,  $SD = 0.22$ ) and low-semantic items [ $M = 0.45$ ,  $SD = 0.23$ , paired two-sided  $t$ -test,  $t(28) = 2.36$ ,  $p = 0.025$ ].

## Discussion

As detailed below, the results of Experiment 3 provided further generalization to our previous experiments, and extend the relevant findings to the temporal domain. From a general metacognitive view, the results of Experiment 3 were quite similar to those of Experiments 1 and 2. The temporal order task produced an overconfidence bias. The discrimination (of confidence judgments) between correct and incorrect temporal order responses (i.e., resolution) was high and better particularly for the high-semantic items, and both item types exhibited high resolution regarding the item's appearance. On the other hand, in this experiment, the appearance calibration was influenced by both the item type and its serial position.

The current findings also replicated the same appearance errors observed in Experiment 2, as participants erred in about 12% of the trials, even though the memory set-size was within



VWM capacity limits and there was no proactive interference. Importantly, low-semantic items were more susceptible to these errors compared to the high-semantic items, especially when these items appeared at the beginning of the memory array. Similar to Experiment 2, these errors were followed with high confidence that the probed item did not appear. This critically points to a gap in subjective judgments' reliability (Bona et al., 2013; Adam and Vogel, 2017), suggesting that participants were "confidently-blind" to their errors regarding whether the probed item appeared or not. In contrast to Experiment 2's results, however, when the item was reported as "did not appear"—the overall temporal order accuracy was not better than chance. Nevertheless, the overall accuracy was lower in Experiment 3 than in Experiment 2 [ $M = 47.7$  vs.  $52.9$ , one-sided independent-samples  $t$ -test,  $t(63) = 2.1$ ,  $p = 0.019$ ] and thus it is possible that this overall reduction in performance can account for the difference between the two experiments.

More importantly, we found that meaning played a significant role in VWM and in the metacognitive processes of VWM. Consistent with previous findings, meaning enhanced VWM performance (Brady et al., 2016; Shoval and Makovski, submitted). There were fewer appearance errors for high-semantic than for low-semantic items. Yet, participants not only better remembered these items, but they were also more confident, exhibited better resolution in the temporal order task, and were better calibrated for the item's appearance when asked about high-semantic items compared to low-semantic items.

## GENERAL DISCUSSION

The present study explored the metacognitive processes in VWM. Unlike other VWM studies, which typically tested simple stimuli (colors, orientation, etc.), the present study used unique (non-repeating) images of real-world objects as ecological stimuli. These stimuli enabled us to minimize the influence of proactive interference and to test subjective judgments with minimum interference from previous trials intrusions. Experiment 1 examined location accuracy and metacognitive measures for a six-item memory array. Experiment 2 used a four-item memory array for a spatial memory task and directly examined the confidence of observers in their memory of item's appearance. Experiment 3 investigated the subjective and objective performance for both the item's identity and temporal order, as well as the role of the item's semantics in memory and metacognitive performance.

From a general metacognitive perspective, we replicated common findings from other cognitive tasks. Namely, participants consistently exhibited an overconfidence bias, along with moderate resolution (Koriat, 2007; Fiedler et al., 2019). Overconfidence seems to be a persistent and typical finding in VWM and other domains (Pallier et al., 2002; Dentakos et al., 2019). As abovementioned, the resolution estimates for meaningful objects were higher than the resolution reported in previous studies (but see, Masson and Rotello, 2009). Taken together, calibration and resolution may dissociate, with

adequate resolution and poor calibration (i.e., overconfidence) as they are the product of different mechanisms (Koriat, 2007).

Moreover, the findings from Experiments 2 and 3, and specifically, the highly confident memory failures—in which participants erroneously reported that the probed item did not appear, further challenge the reliability of subjective judgments. Previous studies showed mixed results in that regard, where in one study observers rarely exhibited memory failures with high confidence (Rademaker et al., 2012), whereas in another study, observers were mostly blind to their memory failures (Adam and Vogel, 2017). The current results are in line with the latter and challenge the reliability of assessments regarding one's own working memory performance—it resulted in a consistent overestimation of performance along with underestimation of memory failures, even within capacity limits and in the absence of intrusions from previous trials. That is, the assessment of VWM content seems to be subjected to biases, such as "blind" errors, and the overestimation of location and temporal memory performance.

It is also noteworthy that while objective accuracy was better for the last appearing items in the spatial, temporal, and identity tasks, participants were well-calibrated for these memory items only in the identity and spatial tasks. However, the robustness of this finding is not very clear because for the spatial task it was driven by only two data points under a high memory-load condition and calibration in the identity tasks is difficult to assess because of the overall high performance in this task. Thus, additional research is still needed in order to establish the cases in which participants are well-calibrated.

Similarly, future research should scrutinize the phenomenon in which observers report that fully visible items did not appear even though VWM was not full (e.g., Chen et al., 2019). This seems to be an intriguing and unexpected effect as one would expect that four real-world, visually distinct objects should be easily remembered. Furthermore, more research is specifically required for clarifying the finding that observers had better than chance knowledge about items that they reported as did-not-appear in the spatial memory task of Experiment 2, but not in the temporal memory task of Experiment 3.

Only a few studies have previously examined how the task requirements (e.g., spatial, temporal, or identity tasks) affect metacognitive measures, as most studies usually rely on the probed-item identity as the objective measure (e.g., "is this color old or new?"). A notable exception is a study employing a metacognitive framework that focused on age differences in a spatial working memory task and found that spatial information resulted in better performance compared to identity information. Furthermore, while identity accuracy decreased as a function of the memory set-size, the location accuracy remained mostly unaffected. However, from a metacognitive perspective, the resolution of identity information was higher than the resolution of location information (Exp. 1a, 1b, Thomas et al., 2012). Given the methodological differences between this and the current study, it is hard to draw any conclusions. But it is also important to note that improved performance may result in higher resolution (Rouault et al., 2018), and eventually identity information is often associated with a specific location (e.g., Toh et al., 2020).

Other studies that examined the metacognitive monitoring in temporal tasks showed mixed results. One study found that observers were accurate in reproducing temporal information and were aware of their errors (Akdogan and Balci, 2017). Consistent with the current results, another study found that participants were largely unaware of their errors and without explicit feedback on errors, participants overestimated their performance in a temporal task (Riemer et al., 2019). Taken together, the current results seem to suggest that the basic metacognitive principles and biases apply regardless of the exact task at hand.

In addition, the findings from Experiment 3 imply that stimuli meaning and semantics play a role in VWM and metacognitive judgments. The results showed that high-semantic items were more accurate and less prone to appearance errors. They were also rated with greater confidence than distorted, low-semantic items. The resolution for high-semantic items in a temporal task was better than for low-semantic items, but both item types exhibited the same degree of false-confidence when participants reported that the probed-item did not appear.

This pattern suggests that observers were able to add a semantic label to an image—one which improved the following recognition of that image (see Brady et al., 2016; Shoval and Makovski, submitted). Meaning could improve VWM performance by different routes. It is possible, for instance, that meaning acts as a conceptual hook: the label of an item essentially adds another level of available information, which can later be used as a “retrieval” cue (e.g., Konkle et al., 2010). Another possibility is that relying on previous knowledge (i.e., item’s semantics) might reduce the amount and complexity of the visual information needed to actively maintain the item’s representation in VWM (because previous knowledge about an item is probably associated with at least some “prototypical” visual features and therefore reduces the information entropy).

This involvement of semantics in VWM could not only ease the maintenance which in turn, allows for better recognition, but could also make the item’s visual representation and associated information more accessible for later judgments (see Berry et al., 2013, for findings that suggest that meaningful memory items improve metacognition). Be that as it may, these findings suggest that the use of real-world objects in VWM tasks lead to the involvement of long-term memory in that task, and thus should be taken into account when trying to isolate VWM capacity measures (Shoval and Makovski, submitted).

In sum, the metacognition of VWM of real-world objects seems to follow a similar pattern of metacognitive results, especially those suggesting overestimation of performance as well as underestimation of “blind” errors. These biases seem not to be task-specific as they were found in temporal, spatial, and identity VWM tasks. We further found that the use of meaningful, complex items, which improved VWM performance, also increased the confidence ratings as well as the metacognitive resolution in a temporal order task. Together these findings challenge the consistency of subjective judgments in VWM and calls for caution in the use of meaningful objects in VWM tasks when attempting to isolate VWM performance.

## DATA AVAILABILITY STATEMENT

The data and tasks supporting this article are available on Open Science Framework (<https://osf.io/nfkj9/>).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Department of Psychology and Education Ethics Committee, The Open University of Israel. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

TM, YS, and TS contributed to the conception and design of the study. YS and TS performed the statistical analysis. TS wrote the first draft of the manuscript. YS and TM provided critical comments on the manuscript. All authors contributed to the manuscript revision, and read and approved the submitted version.

## FUNDING

This study was supported by ISF grant 1344/17 to TM and the Open University of Israel startup grant for YS.

## REFERENCES

- Ackerman, R., and Goldsmith, M. (2011). Metacognitive regulation of text learning: on screen versus on paper. *J. Exp. Psychol.* 17, 18–32. doi: 10.1037/a0022086
- Adam, K. C., Vogel, E. K., and Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cogn. Psychol.* 97, 79–97. doi: 10.1016/j.cogpsych.2017.07.001
- Adam, K. C. S., and Vogel, E. K. (2017). Confident failures: lapses of working memory reveal a metacognitive blind spot. *Atten. Percept. Psychophys.* 79, 1506–1523. doi: 10.3758/s13414-017-1331-8
- Akdogan, B., and Balci, F. (2017). Are you early or late? Temporal error monitoring. *J. Exp. Psychol.* 146, 347–361. doi: 10.1037/xge0000265
- Berry, J. M., Williams, H. L., Usabalieva, A., and Kilb, A. (2013). Metacognitive awareness of the associative deficit for words and names. *Aging Neuropsychol. Cogn.* 20, 592–619. doi: 10.1080/13825585.2012.761670
- Bona, S., Cattaneo, Z., Vecchi, T., Soto, D., and Silvano, J. (2013). Metacognition of visual short-term memory: dissociation between objective and subjective components of VSTM. *Front. Psychol.* 4:62. doi: 10.3389/fpsyg.2013.00062
- Bona, S., and Silvano, J. (2014). Accuracy and confidence of visual short-term memory do not go hand-in-hand: behavioral and neural dissociations. *PLoS One* 9:e90808. doi: 10.1371/journal.pone.0090808
- Brady, T. F., Konkle, T., Alvarez, G. A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14325–14329. doi: 10.1073/pnas.0803390105
- Brady, T. F., Störmer, V. S., and Alvarez, G. A. (2016). Working memory is not fixed-capacity: more active storage capacity for real-world objects than for

- simple stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7459–7464. doi: 10.1073/pnas.1520027113
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–436. doi: 10.1163/156856897x00357
- Broadbent, D. E., and Broadbent, M. H. (1981). Recency effects in visual memory. *Q. J. Exp. Psychol. Sec. A* 33, 1–15. doi: 10.1080/14640748108400762
- Caplovitz, G. P., Fendrich, R., and Hughes, H. C. (2008). Failures to see: attentive blank stares revealed by change blindness. *Conscious. Cogn.* 17, 877–886. doi: 10.1016/j.concog.2007.08.006
- Chen, H., Yu, J., Fu, Y., Zhu, P., Li, W., Zhou, J., et al. (2019). Does attribute amnesia occur with the presentation of complex, meaningful stimuli? The answer is, “it depends”. *Mem. Cogn.* 47, 1133–1144. doi: 10.3758/s13421-019-00923-7
- Dentakos, S., Saoud, W., Ackerman, R., and Toplak, M. E. (2019). Does domain matter? Monitoring accuracy across domains. *Metacogn. Learn.* 14, 413–436. doi: 10.1007/s11409-019-09198-4
- Endress, A. D., and Potter, M. C. (2014). Large capacity temporary visual memory. *J. Exp. Psychol.* 143, 548–565. doi: 10.1037/a0033934
- Fiedler, K., Ackerman, R., and Scarampi, C. (2019). “Metacognition: monitoring and controlling one’s own knowledge, reasoning and decisions,” in *The Psychology of Human Thought: An Introduction*, ed. R. Sternberg (Norderstedt: BoD – Books on Demand), 89–103.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *Am. Psychol.* 34, 906–911. doi: 10.1037/0003-066x.34.10.906
- Fleming, S. M., and Lau, H. C. (2014). How to measure metacognition. *Front. Hum. Neurosci.* 8:443. doi: 10.3389/fnhum.2014.00443
- Goldsmith, M., Pansky, A., and Koriati, A. (2014). “Metacognitive control of memory reporting,” in *The SAGE Handbook of Applied Memory*, eds T. J. Perfect, and D. S. Lindsay (Thousand Oaks, CA: SAGE), 481–500. doi: 10.4135/9781446294703.n27
- Hartshorne, J. K. (2008). Visual working memory capacity and proactive interference. *PLoS One* 3:e2716. doi: 10.1371/journal.pone.0002716
- Hughes, H. C., Caplovitz, G. P., Loucks, R. A., and Fendrich, R. (2012). Attentive and pre-attentive processes in change detection and identification. *PLoS One* 7:e42851. doi: 10.1371/journal.pone.0042851
- Keppel, G., and Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *J. Verb. Learn. Verb. Behav.* 1, 153–161. doi: 10.1016/s0022-5371(62)80023-1
- Köhler, S., Moscovitch, M., and Melo, B. (2001). Episodic memory for object location versus episodic memory for object identity: do they rely on distinct encoding processes? *Mem. Cogn.* 29, 948–959. doi: 10.3758/BF03195757
- Komori, M. (2016). Effects of working memory capacity on metacognitive monitoring: a study of group differences using a listening span test. *Front. Psychol.* 7:285. doi: 10.3389/fpsyg.2016.00285
- Konkle, T., Brady, T., Alvarez, G., and Oliva, A. (2010). Remembering thousands of objects with high fidelity. *J. Vis.* 8, 694–694. doi: 10.1167/8.6.694
- Koriati, A. (2007). “Metacognition and consciousness,” in *The Cambridge Handbook of Consciousness*, eds P. D. Zelazo, M. Moscovitch, and E. Thompson (New-York, NY: Cambridge University Press), 289–325. doi: 10.1017/CBO9780511816789.012
- Lin, P., and Luck, S. J. (2012). Proactive interference does not meaningfully distort visual working memory capacity estimates in the canonical change detection task. *Front. Psychol.* 3:42. doi: 10.3389/fpsyg.2012.00042
- Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual. *Trends Cogn. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006
- Makovski, T. (2016). Does proactive interference play a significant role in visual working memory tasks? *J. Exp. Psychol.* 42, 1664–1672. doi: 10.1037/xlm0000262
- Makovski, T. (2018). Meaning in learning: contextual cueing relies on objects’ visual features and not on objects’ meaning. *Mem. Cogn.* 46, 58–67. doi: 10.3758/s13421-017-0745-9
- Makovski, T., and Jiang, Y. V. (2008). Indirect assessment of visual working memory for simple and complex objects. *Mem. Cogn.* 36, 1132–1143. doi: 10.3758/MC.36.6.1132
- Makovski, T., Watson, L. M., Koutstaal, W., and Jiang, Y. V. (2010). Method matters: systematic effects of testing procedure on visual working memory sensitivity. *J. Exp. Psychol.* 36, 1466–1479. doi: 10.1037/a0020851
- Maniscalco, B., and Lau, H. (2015). Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task. *Neurosci. Conscious.* 2015:niv002. doi: 10.1093/nc/niv002
- Masson, M. E., and Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. *J. Exp. Psychol.* 35, 509–527. doi: 10.1037/a0014876
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychol. Bull.* 95, 109–133. doi: 10.1037//0033-2909.95.1.109
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., et al. (2002). The role of individual differences in the accuracy of confidence judgments. *J. Gen. Psychol.* 129, 257–299. doi: 10.1080/00221300209602099
- Pertzov, Y., Dong, M. Y., Peich, M.-C., and Husain, M., (2012). Forgetting what was where: the fragility of object-location binding. *PLoS One* 7:e48214. doi: 10.1371/journal.pone.0048214
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team, (2019). *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-143*. Available at: <https://CRAN.R-project.org/package=nlme> (accessed October 09, 2019).
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Available at: <https://www.r-project.org/>
- Rademaker, R. L., Tredway, C. H., and Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *J. Vis.* 12, 21–21. doi: 10.1167/12.13.21
- Riemer, M., Kubik, V., and Wolbers, T. (2019). The effect of feedback on temporal error monitoring and timing behavior. *Behav. Brain Res.* 369:111929. doi: 10.1016/j.bbr.2019.111929
- Rouault, M., McWilliams, A., Allen, M. G., and Fleming, S. M. (2018). Human metacognition across domains: insights from individual differences and neuroimaging. *Pers. Neurosci.* 1:e17. doi: 10.1017/pen.2018.16
- Samaha, J., and Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proc. R. Soc. B* 284:1867. doi: 10.1101/140558
- Schwartz, B. L., and Eklides, A. (2012). Metamemory and memory efficiency: implications for student learning. *J. Appl. Res. Mem. Cogn.* 1, 145–151. doi: 10.1016/j.jarmac.2012.06.002
- Shoval, R., Luria, R., and Makovski, T. (2019). Bridging the gap between temporary memory and working memory: the role of stimuli distinctiveness. *J. Exp. Psychol.* doi: 10.1037/xlm0000778 [Epub ahead of print].
- Suchow, J. W., Fougner, D., and Alvarez, G. A. (2017). Looking inward and back: real-time monitoring of visual working memories. *J. Exp. Psychol.* 43, 660–668. doi: 10.1037/xlm0000320
- Thomas, A. K., Bonura, B. M., Taylor, H. A., and Brunyé, T. T. (2012). Metacognitive monitoring in visuospatial working memory. *Psychol. Aging* 27, 1099–1110. doi: 10.1037/a0028556
- Toh, Y. N., Sisk, C. A., and Jiang, Y. V. (2020). Effects of changing object identity on location working memory. *Attent. Percept. Psychophys.* 294–311. doi: 10.3758/s13414-019-01738-z
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Vandenbroucke, A. R., Sligte, I. G., Barrett, A. B., Seth, A. K., Fahrenfort, J. J., and Lamme, V. A. (2014). Accurate metacognition for visual sensory memory representations. *Psychol. Sci.* 25, 861–873. doi: 10.1177/0956797613516146
- Vlassova, A., Donkin, C., and Pearson, J. (2014). Unconscious information changes decision accuracy but not confidence. *Acad. Sci. U.S.A.* 111, 16214–16218. doi: 10.1073/pnas.1403619111
- Yue, C. L., Castel, A. D., and Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: the influence of typeface clarity on metacognitive judgments and memory. *Mem. Cogn.* 41, 229–241. doi: 10.3758/s13421-012-0255-8

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sahar, Sidi and Makovski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Visual Working Memory of Chinese Characters and Expertise: The Expert's Memory Advantage Is Based on Long-Term Knowledge of Visual Word Forms

Hubert D. Zimmer<sup>1\*</sup> and Benjamin Fischer<sup>2</sup>

<sup>1</sup> Brain & Cognition Unit, Department of Psychology, Saarland University, Saarbrücken, Germany, <sup>2</sup> International Research Training Group "Adaptive Minds", Saarland University, Saarbrücken, Germany

## OPEN ACCESS

### Edited by:

Marius Peelen,  
Radboud University, Netherlands

### Reviewed by:

Weizhen Xie,  
National Institutes of Health (NIH),  
United States  
Alan C.-N. Wong,  
The Chinese University of Hong Kong,  
China

### \*Correspondence:

Hubert D. Zimmer  
huzimmer@mx.uni-saarland.de

### Specialty section:

This article was submitted to  
Cognition,  
a section of the journal  
Frontiers in Psychology

**Received:** 29 August 2019

**Accepted:** 04 March 2020

**Published:** 17 April 2020

### Citation:

Zimmer HD and Fischer B (2020)  
Visual Working Memory of Chinese  
Characters and Expertise:  
The Expert's Memory Advantage Is  
Based on Long-Term Knowledge  
of Visual Word Forms.  
Front. Psychol. 11:516.  
doi: 10.3389/fpsyg.2020.00516

People unfamiliar with Chinese characters show poorer visual working memory (VWM) performance for Chinese characters than do literates in Chinese. In a series of experiments, we investigated the reasons for this expertise advantage. Experiments 1 and 2 showed that the advantage of Chinese literates does not transfer to novel material. Experts had similar resolution as novices for material outside of their field of expertise, and the memory of novices and experts did not differ when detecting a big change, e.g., when a character's color was changed. Memorizing appears to function as a rather abstract representation of word forms because memory for characters' fonts was poor independently of expertise (Experiment 3), though still visual. Distractors that were highly similar conceptually did not increase memory errors, but visually similar distractors impaired memory (Experiment 4). We hypothesized that literates in Chinese represent characters in VWM as tokens of visual word forms made available by long-term memory. In Experiment 5, we provided novices with visual word form knowledge. Participants subsequently performed a change detection task with trained and novel characters in a functional magnetic resonance experiment. We analyzed set size- and training-dependent effects in the intraparietal sulcus (IPS) and the visual word form area. VWM for trained characters was better than for novel characters. Neural activity increased with set size and at a slower rate for trained than for novel characters. All conditions approached the same maximum, but novel characters reached the maximum at a smaller set size than trained characters. The time course of the bold response depended on set size and knowledge status. Starting from the same initial maximum, neural activity at small set sizes returned to baseline more quickly for trained characters than for novel characters. Additionally, high performers showed generally more neural activity in the IPS than low performers. We conclude that experts' better performance in working memory (WM) is caused by the availability of visual long-term representations (word form types) that allow a sparse representation of the perceived stimuli and make even small changes big because they cause a type change that is easily detected.

**Keywords:** visual working and short-term memory, visual working memory c, expertise, visual working memory precision, Chinese character, visual word form area



## INTRODUCTION

Visual working memory enables the storage of about three to four visual objects for a short period of time (Luck, 2008; Zimmer, 2008; Luck and Vogel, 2013; Mance and Vogel, 2013). It is therefore assumed that VWM has a highly limited capacity, where “capacity” denotes the number of items or features that can on average be remembered. Capacity is commonly estimated by the performance in a so-called change detection task. A study picture displaying a varying number of objects is briefly presented, and after a short interval of about 1 s, a test display is shown. The object(s) in the test display are either the same as those in the study picture or presented with some changes. The participant’s task is to detect these changes. The classical result is that performance is nearly perfect up to a set size of three to four items and subsequently decreases sharply (Luck and Vogel, 1997). When specific assumptions about the decision process are made, the proportion of changes identified correctly can be used to estimate capacity as the number of items that can be represented in working memory (WM). Across many domains, it is around four (Cowan, 2001). This capacity is smaller if the items to be memorized are visually complex (Alvarez and Cavanagh, 2004). For Chinese characters, the capacity of VWM can be as low as one character (Zimmer and Fu, 2008; Sun et al., 2011; Schurgin and Brady, 2019). However, this effect varies strongly with the expertise of the participants. Literate Chinese show better memory for familiar than unfamiliar characters (e.g., Yu et al., 1985; Zhang and Simon, 1985), and experts have a higher capacity for Chinese characters than novices who are inexperienced with this material (Hue and Erickson, 1988; Sun et al., 2011).

The aim of this paper is to investigate the mechanism of this expertise advantage in VWM. In a series of experiments, we will show that knowledge of visual word forms is the key factor. We assume that this allows experts to represent characters as sparse code at a structurally high level. We consider the VWM entry as a token of the word form representation (the character’s orthography) in long-term memory. By this mechanism, the percepts of complex visual characters are reduced to token representations of known items. In change detection tasks, such representations allow an item change to be detected easily because the change causes a shift to a different “visual category” (a different character), and it does so even if the change is perceptually small. Oberauer and Eichenberger (2013) speculated that “the unit of VWM is the largest unified representation in long-term memory that can be used to code the memoranda” (p. 1219). For VWM of Chinese characters, our data suggest that long-term knowledge of visual word forms provides this code and that long-term memory, shaped by a participant’s experience, contributes to working memory.

That experts show better working memory than novices has also been shown in other domains. For example, people show better working memory for upright faces with which we are all familiar (i.e., expert) than for inverted faces (Curby and Gauthier, 2007). Similarly, when working memory for cars was tested, car experts showed better memory than car novices, but this advantage disappeared when the cars were presented upside-down (Curby et al., 2009). VWM for famous faces is

also better than for unfamiliar faces, and again, the advantage disappears for inverted faces, i.e., in an unfamiliar viewing condition (Jackson and Raymond, 2008). Working memory is also higher for common objects than for color patches if encoding time is long (1 or 2 s) (Brady et al., 2016). Similarly, participants who were familiar with Pokémons memorized more items than people who were unfamiliar with the specific material (Xie and Zhang, 2017b). These results suggest that familiarity with the to-be-memorized material or pre-experimental knowledge of the to-be-memorized items boosts working memory performance. It follows that increasing experience with a specific material should increase working memory performance (Sorensen and Kyllingsbaek, 2012). However, the mechanisms that cause this advantage are controversial.

For example, the mechanism may be a variant of chunking (Simon, 1974). The features of known items can be chunked and represented as one unit, whereas unknown items are represented in smaller pieces. If one assumes that VWM provides a small number of “slots,” a novel item would fill more slots than a familiar one. If the study array depicts four items, working memory may be capable of representing three of these items if they are familiar. However, the capability may be reduced to one item if the items are novel. If a complex novel character is decomposed into part figures and each sub-figure is individualized, each would fill one “slot.” In this way, one item counts functionally as two or three. A consequence is that the other items in the display cannot be represented and are lost.

Barton et al. (2009) argued against this interpretation. They claimed that working memory always represents “objects” as units and that individual features of any object are not distributed to more than one slot. Hence, even when novel items are seen, working memory should represent about three of them. The authors suggest that poor performance may be a consequence of errors in the comparison process. With complex material such as faces and Chinese characters, changes may often be missed because they are small. Should the change be large (e.g., if a character were changed into a cube), it would be detected, and performance would improve, showing memory for about three items. This is exactly what the authors observed (Awh et al., 2007). The authors therefore distinguished between the number of objects that are represented in working memory and the precision or resolution of these representations. The former was estimated from the detection rate of large changes, and this number correlated with capacity estimations gained from simple material, e.g., color patches (Awh et al., 2007). The same result was observed when comparing working memory for upright and inverted faces (Scolari et al., 2008). The authors took this as evidence that expertise enhances the precision of working memory representation but not capacity in terms of the number of “slots.” The number of items represented should be independent of expertise, but unfamiliar items are represented less precisely than familiar ones (Lorenc et al., 2014).

Even though it is highly meaningful that comparison errors are a function of visual similarity between the study item and test item, some authors have argued against this interpretation of the results. Morey et al. (2012) put forward theoretical arguments. They criticized the computation of the precision parameter

and argued that it can only be taken as a measure of the individually experienced distractor similarity, not as a quality of working memory. Other counter-arguments were empirical. The probability of detecting a large change does not only depend on the similarity between the study item and test item but also on the perceptual features of the context in which the study item is presented (ensemble representations) (Brady et al., 2019). This controversy makes it obvious that it is not sufficient to speak of capacity in terms of slots. The similarity between a study item and a test item depends on what is represented in memory, and this is a function of the perceiver's expertise. Therefore, the perceived similarity, not the physical similarity, is what is relevant to the comparison (Zimmer, 2008; Dall and Sorensen, 2019).

This brings another mechanism to our attention. Expertise influences the quality of perceptual processing. In the field of expertise, stimuli may be encoded by comparisons with templates in long-term memory. With characters, these are word form entries (Dehaene et al., 2005), which are activated via radicals (Perfetti et al., 2005; Chen and Yeh, 2015). Experts can represent a character as a perceptual token of a word form type. This token is constructed by interactive activation of perceptual output and representations of the orthographic word forms (types) in long-term memory (Leck et al., 1995). In the bottom-up stream, orthographic features, e.g., the number of strokes (Sze et al., 2015), influence character encoding, whereas qualities of word forms, e.g., character frequency or diagnostic features of stroke patterns, influence perception in the top-down stream. As a consequence, Chinese literates may need less perceptual support for identification than novices and fill in information that they missed. In contrast, Chinese illiterates cannot make use of word form knowledge. For these people, processing is oriented to the stroke level (Yeh et al., 2003). They can represent items based only on the bottom-up stream, which makes perception more demanding and often causes incomplete representations that cannot be compensated by long-term knowledge. One consequence is that novices show poorer memory for visually complex characters than for simple characters, a factor that is less relevant for Chinese literates (Zimmer and Fu, 2008; Sun et al., 2011).

Collectively, these data suggest that word form knowledge makes an important contribution to character encoding, and we hypothesize that it also adds to the expertise advantage in VWM. Initial support for this claim is the observation that acquiring long-term knowledge of characters enhances VWM for these characters compared to novel characters and changes neural activity in the visual word form area (VWFA) (Zimmer et al., 2012). The series of experiments reported in this paper should provide further evidence for this word form token hypothesis and the argument that the advantage of Chinese literates in VWM for Chinese characters is a consequence of word form knowledge in long-term memory.

A third route by which expertise can influence VWM performance is a difference in low-level visual processing. The visual processing demands of Chinese characters are high, giving orthography a special relevance (Perfetti et al., 2013). Characters often consist of many strokes, and increasing the number of strokes increases a character's complexity. Perimetric

complexity—a geometrically defined measure of a figure's complexity—is clearly larger for Chinese characters than for standard fonts of Latin alphabets (Ngiam et al., 2019). In a study by Ma and Chuang (2015), characters with many strokes or stroke crossings as well as enclosed characters were judged as subjectively the most complicated. In another study, complexity—defined by the number of strokes—had similar effects on search times in experts and novices (Yu et al., 2018), which suggests that complexity has an early effect in visual processing. If perception of complex characters is a demanding task, it is conceivable that extended practice in identifying such visually dense stimuli provides Chinese literates with enhanced perceptual skills (McBride and Wang, 2015). Some results speak in favor of this possibility. Visual-spatial abilities were found to be positively correlated with the learning of Chinese characters (McBride-Chang et al., 2005). Additionally, learning to read enhances perceptual skills (McBride-Chang et al., 2011). In a cross-cultural study, it was observed that Chinese scholars exclusively outperformed Greek scholars in visual/spatial processing tasks (Demetriou et al., 2005). As yet, however, we cannot provide evidence for this knowledge transfer hypothesis. In our previous studies, Chinese literates did not show an advantage in VWM for other types of complex visual material that were structurally similar to Chinese characters (e.g., Zimmer and Fu, 2008; Sun et al., 2011).

Although evidence for the transfer of encoding skills to novel visual material is lacking, it is still possible that experts process items within their field of expertise differently compared to novices. For example, experts might focus specifically on diagnostic features. In Wagar and Dixon's (2005) study, for instance, Greeble experts detected changes of diagnostic features more frequently and non-diagnostic features less frequently than novices, who recognized both changes equally often. Another possibility is that experts represent more high spatial frequencies compared to novices. It has been suggested that items are initially represented in working memory at a coarse level and more detailed information is then added (Gao et al., 2009, 2010; Gaspar et al., 2013). According to such models, the aggregation of information over time proceeds from global shape to fine details (Alvarez and Cavanagh, 2008; Xu and Chun, 2009). Because the availability of long-term knowledge influences the speed of this process (Brady et al., 2009), it may also change the level (the resolution) that is finally reached if known visual material is encoded.

Finally, one can think of other, non-visual influences of expertise on working memory. For example, for Chinese literates, words have phonology and meaning. However, even if they are available, phonological codes do not seem to be relevant in VWM tasks. The presence or absence of articulatory suppression was found to have no influence on VWM performance (Morey and Cowan, 2004; Jackson and Raymond, 2008; Sun et al., 2011; Xie and Zhang, 2017c; Ngiam et al., 2019). It is probable that timing and the large number of trials do not favor a naming strategy. However, articulation impairs memory if it has side effects on visual processes, e.g., if words can be imagined (Mate et al., 2012). Also, a verbal preload does not change the working memory effects (Luck and Vogel, 1997;

Vogel et al., 2001; Wagar and Dixon, 2005). Therefore, we do not believe that characters are phonologically recoded. The usage of semantic codes will be discussed in Experiment 4.

Obviously, several non-exclusive mechanisms can contribute to the expertise advantage in temporary memory for Chinese characters. In the following experiments, we investigate the relevance of the different components. Experiments 1 to 4 involved change detection tasks with different types of material. Experiment 5 consisted of a training study in which we taught a set of characters and compared the VWM for trained and novel characters.

## GENERAL PROCEDURE

All participants were either native Chinese speakers (experts) or Germans with no knowledge of Chinese language (novices). In Experiment 5, participants were native German speakers. The trial structure matched the standard procedure of a change detection task. A study display was presented for a short period. It depicted a variable number (set size) of “objects” (characters, figures, color patches). After a short retention interval, a single “object” was shown, and participants were required to decide whether this item had been presented during the study or something had changed. The experiments were controlled by E-Prime 2. The details and material of the experiments are presented in the **Supplementary Material** to this paper. The studies were approved by the ethics committee of the Fakultät für Empirische Humanwissenschaft, Saarland University, and all participants gave their informed consent.

We counted the proportions identified correctly and calculated corrected recognition scores ( $Pr$ ) for each condition.  $Pr$  is the difference between the proportion of hits (detected changes) in each change condition and the proportion of false alarms (change responses to unchanged items) to all items in the same stimulus category, e.g., all unchanged characters or color patches. The  $Pr$  data were analyzed by repeated measurement analyses of variance (ANOVA). If sphericity was violated, a Huynh-Feldt correction was applied and epsilon reported.

## EXPERIMENT 1

The first experiment tested whether novices store fewer items of a study display compared to experts. If a complex item “consumes” full capacity, novices should represent only one item at the expense of other displayed items. For example, Chinese illiterates might represent parts of one item in three “slots” and fully ignore other items. However, it is also possible that novices represent about three items similarly to Chinese literates, but only one might embody details, while the other items are represented merely by coarse information, which is rarely appropriate to detect a character change. Should coarse information be sufficient, the change would be detected in more than one item. In order to show this, it is necessary to measure memory performance for Chinese characters with “big” changes. We presented characters in color ink and then changed the color

of an item from study to test. If Chinese illiterates store only one item of the study array, they would miss even the color change in a character because the “object” that had changed color was not represented in working memory at all. In contrast, if the low performance for characters is a matter of “resolution,” a color change (big change) would be detected, but a character change (small change) would not be detected. To test this, we ran a change detection experiment with colored Chinese characters in which the character, its color, or both could change. For controls, color patches were presented in a fourth condition.

## Participants

Twenty-eight Chinese students were tested in Beijing, and 20 German students from Saarland University also took part in the experiment. One Chinese participant was excluded due to below-chance performance.

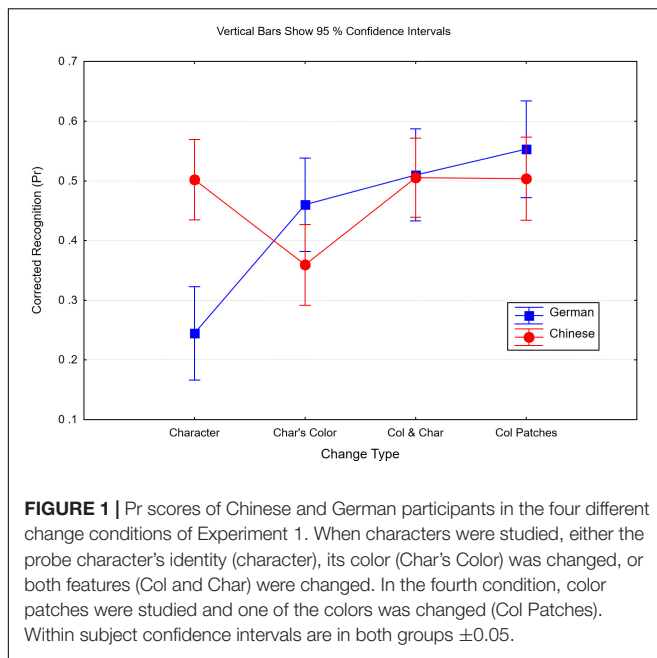
## Design and Procedure

The study was a  $2 \times 4$  mixed design with expertise (Chinese, German) and change type as the factors. Change type included four levels: three with character arrays as study material (the character changed, character’s color changed, both features changed) and one with color patches (color changed). The set size was four, the study time 500 ms, and the retention interval 1,000 ms.

## Results and Discussion

We calculated  $Pr$ , as shown in **Figure 1**. Having calculated the proportion correctly identified in the different change conditions, we subtracted from each accuracy rate the rate of change responses to no-change trials (false alarms) with the same item material, i.e., characters or color patches. In an ANOVA with expertise (Chinese, German participants) and change type as factors, the two factors interacted [ $F(3,135) = 23.10$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.34$ , mean standard error ( $MSE$ ) = 0.013,  $\epsilon = 0.88$ ]. In both groups, the simple main effects were significant [German:  $F(3,57) = 37.12$ ,  $p < 0.0001$ ,  $\epsilon = 0.73$ ; Chinese:  $F(3,78) = 9.69$ ,  $p < 0.0001$ ,  $\epsilon = 0.76$ ]. However, the profiles of performance over change types were different. Germans showed poor memory only when characters changed (small change) but not when their color changed (big change). However, in one of the conditions in which color was changed, the two groups differed. Chinese detected the changed color of an unchanged character worse ( $\Delta = 0.15$ ) than any other change ( $p < 0.0002$  in pairwise comparisons). All other conditions did not differ from each other ( $F < 1$ ).

We inferred that novices represent nearly the same number of “objects” if they encode Chinese characters as they do when they encode visually less complex material (color patches). Therefore, the poor character memory of novices was not due to a general failure of encoding the complex items, i.e., “missed objects.” More likely, at a coarse level, all participants encode as many items as they can individualize (Xu and Chun, 2009), but novices represent fine details of one item only or thereabouts, whereas experts do this for nearly all items of the study display. This is not a consequence of the encoding time being too short for the unfamiliar characters. The encoding rate for unfamiliar characters has been estimated at 12 items per



second (Ngiam et al., 2019), and in a study with variable and masked presentation time, performance increased with time but reached an asymptote at 450 ms (Sun et al., 2011). A study time of 500 ms should therefore be long enough to encode four items.

## EXPERIMENT 2

If the number of represented objects is no different but experts represent “objects” in more detail, this may be due to a general competence or may be restricted to the field of expertise. Their long experience with complex visual input may have enhanced the perceptual skills of Chinese literates when processing fine details in general. In this case, we should see better memory also for other types of visual material if small changes are to be detected.

Addressing a similar question, Fukuda et al. (2010) used a change detection task with big changes (to shape) or small changes (to the inner patterns of shapes). They constructed figures of different shapes (oval, rectangle, and so on) with differing inner patterns (e.g., two crossing versus two parallel lines). In change trials, an altered outline established a big change, whereas a changed inner pattern was a small change. Memory performance in big-change trials estimated “capacity,” while performance in small-change trials estimated “precision.” In Experiment 2, we used the same manipulation. If Chinese literates have generally enhanced perceptual skills for fine details, they should have better memory for the inner patterns of shapes compared to Germans. With big changes, we do not expect any memory differences.

Additionally, we wanted to replicate the results of Experiment 1. For that purpose, we presented characters in different colors as in Experiment 1, but in Experiment 2, they were also presented in different fonts. Each of these features could change. If Chinese

literates represent perceptual details of characters, we should see a memory advantage for experts even if a character's font was changed. If, however, Chinese literates represent the characters as tokens of visual word forms, font memory should be poor, though the characters' identities should be remembered.

## Participants

Forty-four students from Saarland University took part in the experiment. Twenty-four were native Chinese speakers who studied at Saarland University, and the remainder were native German speakers with no experience of Chinese language.

## Design and Procedure

On the study display, we presented four “objects” (geometrical figures, color patches, or characters) for 500 ms. For the test, one “item” was shown. Figures and color patches defined a  $2 \times 3$  design, with expertise (Chinese, German) and change type (shape, pattern, color patch) as factors. Characters also defined a  $2 \times 3$  design, with expertise (Chinese, German) and change type (color, character, font) as factors. All six trial types were mixed and presented in random order. In change trials, only one feature was altered.

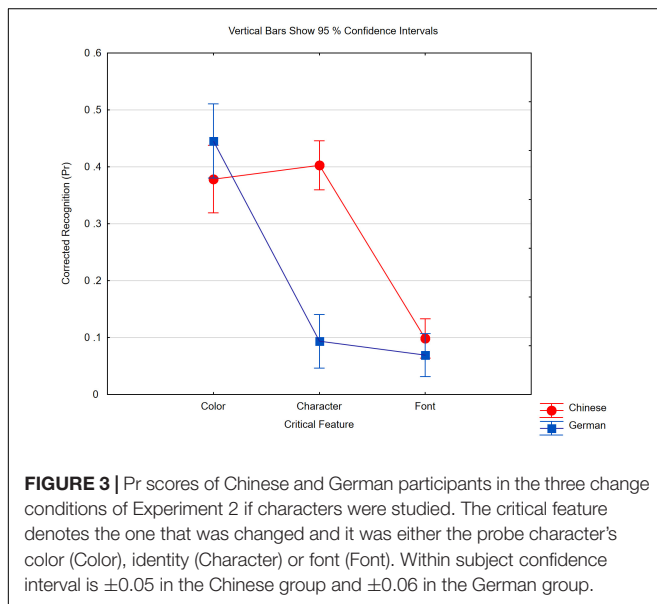
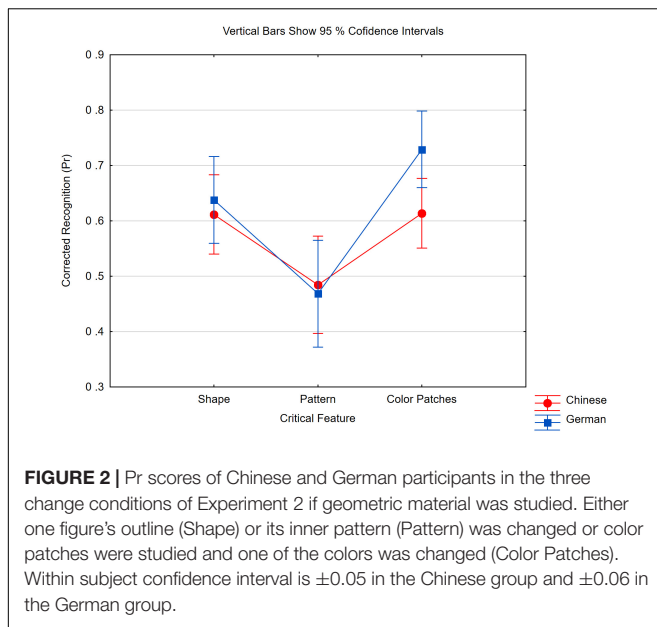
## Results and Discussion

We analyzed the Pr scores in two  $2 \times 3$  ANOVAs with expertise (Chinese, German) and critical feature (shape, pattern, color patch; or color, character, font) as factors. As shown in **Figure 2**, with geometrical material, we obtained a clear effect of the changed feature [ $F(2,84) = 31.11$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.43$ ,  $MSE = 0.015$ ] and an interaction with expertise [ $F(2,84) = 3.37$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.07$ ,  $MSE = 0.015$ ]. The interaction was due to the fact that Germans showed better working memory for color patches than the Chinese did [ $t(42) = 2.45$ ,  $p < 0.05$ ]. Because a difference in color memory was not observed in other experiments from our lab (Zimmer and Fu, 2008; Sun et al., 2011), we considered this result as accidental. When we excluded the color patch condition, the interaction disappeared ( $F < 1$ ). Both groups detected pattern changes less well than shape changes [ $F(1,42) = 30.62$ ,  $p < 0.0001$ ,  $MSE = 0.016$ ,  $\eta_p^2 = 0.42$ ]. Hence, Chinese literates did not show an advantage in detecting small changes to figures.

As shown in **Figure 3**, in the character conditions, we obtained a very strong effect of the changed feature [ $F(2,84) = 112.75$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.73$ ,  $MSE = 0.010$ ] and an interaction with expertise [ $F(2,84) = 40.04$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.49$ ,  $MSE = 0.010$ ]. The interaction was driven by two influences. First, the two groups did not differ if a change to a character's color had to be detected [ $t(42) = 1.53$ ,  $p = 0.13$ ], but Chinese participants detected a character change much better than German participants [ $t(42) = 9.79$ ,  $p < 0.0001$ ]. This replicates the results of Experiment 1. However, both groups did not differ in being very poor at detecting changes to a character's font [ $t(42) = 1.15$ ,  $p = 0.25$ ], even though Chinese participants (but not Germans) remembered the characters' identities.

Obviously, literates in Chinese represented abstract word forms but no detailed visual information on the presented characters. One might argue that this was only strategic: they





habitually focused on a character's shape and ignored its font. In that case, however, it would be surprising that characters' colors were remembered. Nevertheless, to exclude the possibility that we unintentionally missed the experts' memory advantage for fonts, we tested memory for fonts again in Experiment 3, but this time, processing was explicitly focused on the relevant feature.

## EXPERIMENT 3

### Participants

Forty students from Saarland University took part in the experiment. Half were native Chinese speakers studying in Germany, and half were native German speakers.

## Design and Procedure

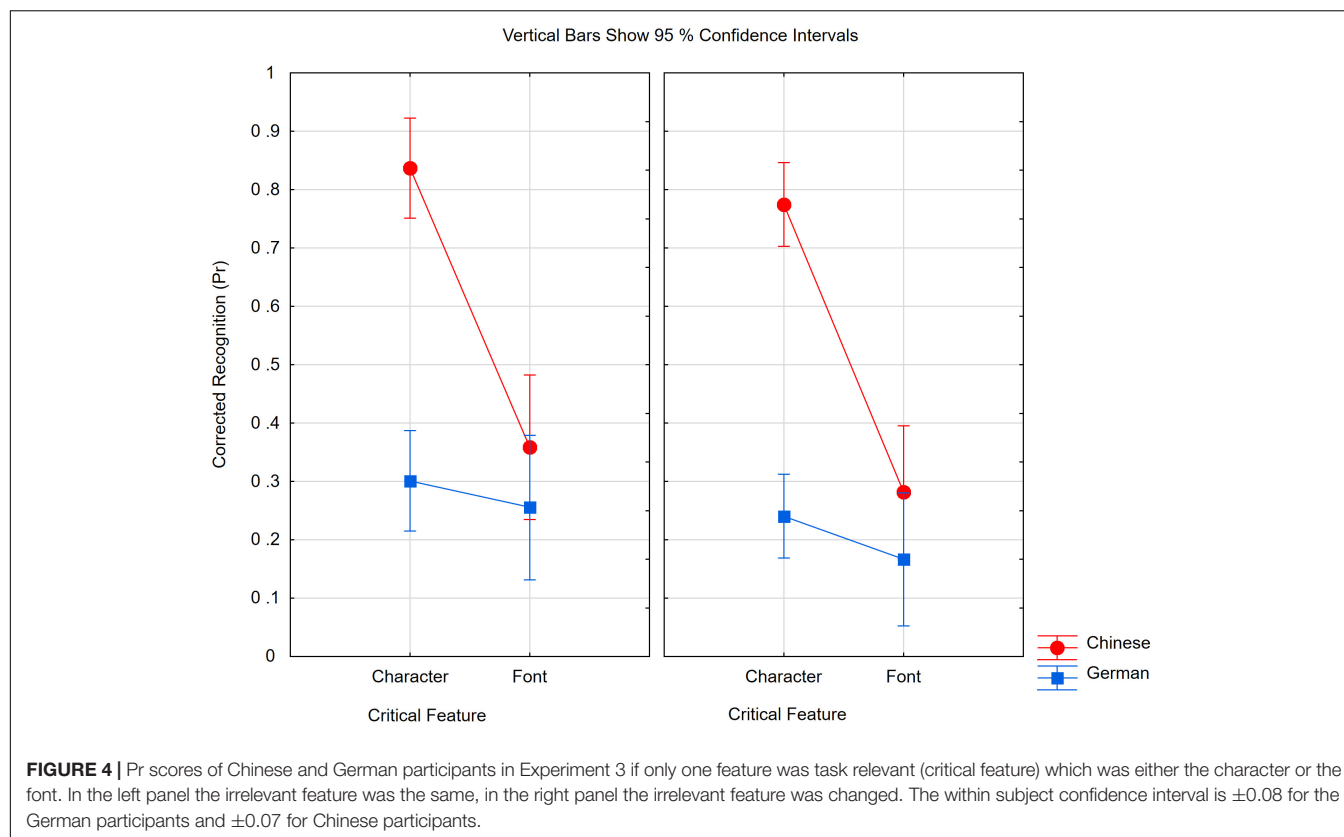
On the study display, three characters were visible for 500 ms (for details, see the **Supplementary Material**). The participants' tasks changed block-wise every 36 trials. In a block, they had to spot a change to a character or a font, or they had to spot both features. We therefore had four types of stimulus matches between study and test: (a) no change; (b) the character was changed but the font was the same; (c) vice versa; and (d) both features were changed. Because each feature changed half of the time irrespective of the task, we could also test irrelevant change effects. If participants could focus successfully on the relevant feature, the match of the irrelevant feature should not matter.

## Results and Discussion

Corrected recognition scores were calculated by subtracting the proportion of false alarms (change decisions to no-change items) from the proportion of detected changes at the same level of irrelevant feature and task. We first analyzed the trials in which only one feature was relevant. An ANOVA was performed with expertise (Chinese, German), change type (character, font), and status of irrelevant feature (same, changed) as the factors. The results of Experiment 2 were fully replicated, as shown in **Figure 4**. We obtained a highly significant interaction between expertise and change type [ $F(1,38) = 46.17, p < 0.0001, \eta_p^2 = 0.55, MSE = 0.039$ ]. Chinese literates detected character changes (0.81,  $SE = 0.04$ ) much better than Chinese illiterates (0.27,  $SE = 0.06$ ). In contrast, the experts' memory for fonts (0.32,  $SE = 0.05$ ) was not significantly better than the novices' memory (0.21,  $SE = 0.05$ ) [ $F(1,38) = 2.19, p = 0.15, \eta_p^2 = 0.05, MSE = 0.030$ ]. The status of the irrelevant feature did not interact with any other factor ( $F < 1$ ), but decisions were more accurate for unchanged (0.44,  $SE = 0.03$ ) than for changed (0.37) irrelevant features [ $F(1,38) = 6.37, p < 0.05, \eta_p^2 = 0.14, MSE = 0.033$ ].

We then analyzed the condition in which both features had to be attended. **Figure 5** depicts the respective data as a function of change type. In an analysis with expertise (Chinese, German) and change type (character, font, both) as factors, we obtained a strong interaction [ $F(2,76) = 32.81, p < 0.0001, \eta_p^2 = 0.46, MSE = 0.020$ ]. The better working memory performance of Chinese literates was confined to conditions in which the character changed [ $F(1,38) = 23.77, p < 0.0001$ ]. If only the font was changed, Chinese participants were even worse than the Germans [ $t(38) = 3.01, p < 0.005$ ]. However, within the German group, the change type had an effect [ $F(2,38) = 6.66, p < 0.005, \epsilon = 0.997$ ]. A Bonferroni test revealed that memory was better if both features had changed (0.41,  $SE = 0.05$ ) than if only one had changed, but memory for characters (0.34,  $SE = 0.03$ ) and fonts (0.26,  $SE = 0.04$ ) did not differ.

Even though we highlighted the relevance of fonts and the test conditions were blocked, we did not find evidence for enhanced memory for fonts among the experts. In both feature blocks, Chinese showed even worse memory performance if only a character's font had changed. This is plausible if we assume that both groups have poor memory for fonts but that, in contrast to Germans, Chinese people recognized the character as the same. A consequence of



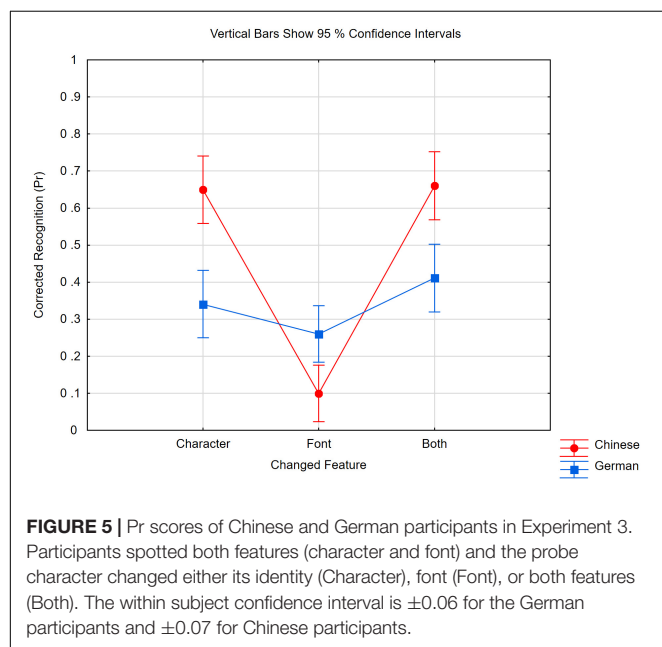
this would be the erroneous no-change decisions. We concluded from these results that expertise did not enhance the representation of low-level details: what was easily detected by Chinese literates (the experts) was the mismatch of the word form.

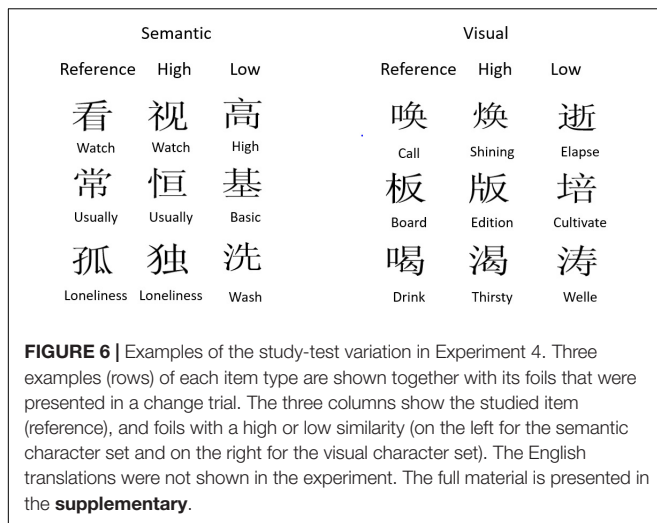
## EXPERIMENT 4

The results of Experiment 4 should provide arguments against an alternative interpretation of the data. Experts have additional codes available that do not exist for novices and that have the potential to support memory. They can name the “objects” and use a phonological code for storage. The observation that articulatory suppression does not change the results (e.g., Curby and Gauthier, 2007; Sun et al., 2011; Sense et al., 2017) is an argument against a naming strategy. However, for Chinese literates, characters also have meaning, and meaning may be available even under articulatory suppression (Bourassa and Besner, 1994; Shivde and Anderson, 2011; Baier and Ansorge, 2019). The results of Experiment 4 provide evidence that Chinese literates still use VWM and not the meaning of the characters.

To test this hypothesis, we manipulated the visual and semantic similarities between the studied and the changed character. Memory performance should be poor if the similarity between the item held in memory and the presented foil is high. If the participant uses a visual memory code, visually similar foils should be erroneously considered as old (Chen et al., 1995; Leck et al., 1995). If the code is semantic, semantically similar items should cause poor performance because the distractor has the same meaning as the study item. As a test, therefore, we compared change detection performance with characters under high and low visual or semantic distractor similarity.

Additionally, we presented pseudo characters as study material. Pseudo words combined two radicals of different real





characters. These “artificial characters” do not exist in Chinese language and therefore have no meaning. This allowed us to manipulate the visual similarity in the change condition in the absence of meaning. We did so by replacing one of the radicals of a studied pseudo word with a visually similar or dissimilar radical but still generating pseudo words. If memory in the visually similar condition is impaired to a similar extent for characters and pseudo characters, this would suggest that working memory for characters is based mainly on visual information.

## Participants

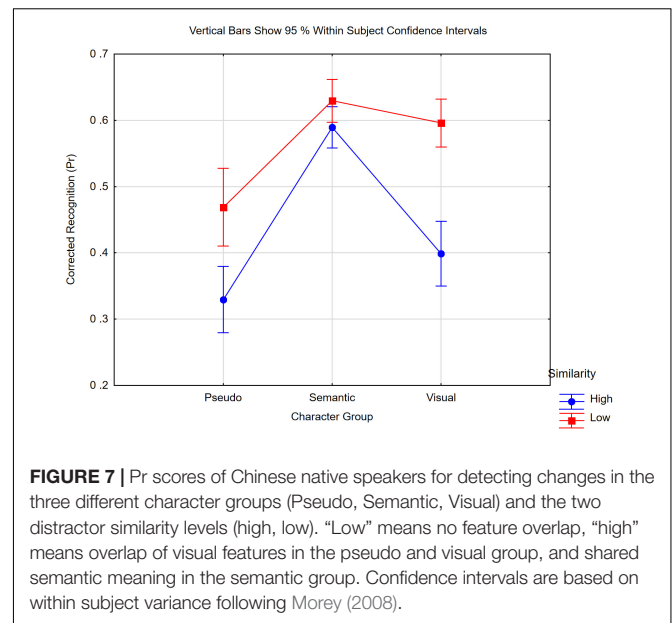
Twenty students from universities in Beijing took part in this experiment.

## Design and Procedure

Two sets of 10 characters were selected, defined by the type of similarity (as shown in the **Supplementary Material**). In the visual set, for each study item, we selected a foil that had a different meaning but a shared radical; in the semantic set, the changed item was a synonym of the study item without shared visual features (see **Figure 6** for examples). Also, highly similar foils to pseudo words shared a radical with the study item, whereas the dissimilar items did not share features with the study item. In order to minimize strategic influences, presentation time was only 200 ms, and an articulatory suppression task was applied (see the **Supplementary Material**).

## Results and Discussion

We calculated for each change type corrected recognition scores, as shown in **Figure 7**. A  $3 \times 2$  repeated measurement ANOVA was conducted with character set (pseudo, semantic, visual) and similarity level (high, low) as the factors. The results showed a strong interaction [ $F(2,38) = 11.36$ ,  $p < 0.0005$ ,  $\eta_p^2 = 0.37$ ,  $MSE = 0.005$ ]. Visual similarity (pseudo: 0.17; characters: 0.20) impaired memory much more than semantic similarity [0.04;  $t(19) = 4.35$ ,  $p < 0.0005$ ], though the latter effect was nearly significant [ $t(20) = 2.06$ ,  $p < 0.06$ ].



In a follow-up, we analyzed the character conditions in a  $2 \times 2$  ANOVA with type of similarity (semantic, visual) and similarity level (high, low) as the factors. Again, the interaction was significant [ $F(1,19) = 19.13$ ,  $p < 0.0005$ ,  $\eta_p^2 = 0.50$ ,  $MSE = 0.007$ ]. Visually similar foils showed the worst memory compared to all other conditions ( $p < 0.0001$ ). When we compared only the visual conditions in a  $2 \times 2$  analysis with word form (pseudo, characters) and similarity level (high, low) as factors, the main effects of word form [ $F(1,19) = 8.13$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.30$ ,  $MSE = 0.024$ ] and visual similarity [ $F(1,19) = 49.35$ ,  $p < 0.0001$ ,  $\eta_p^2 = 0.72$ ,  $MSE = 0.012$ ] were both significant, but the interaction was not significant [ $F(1,19) = 3.11$ ,  $p = 0.10$ ,  $\eta_p^2 = 0.15$ ].

Pseudo characters were memorized worse than real characters, and both were strongly and to the same extent influenced by the visual similarity of the foils. Semantic similarity had only a weak effect. This speaks in favor of the assumption that even Chinese literates represent the items in a change detection task as visual code in working memory.

## EXPERIMENT 5

Collectively, the above experiments provide evidence for the word form token hypothesis, which assumes that characters are represented in VWM as tokens of word forms stored in long-term memory. In this experiment, we investigated neuropsychological evidence for this claim by demonstrating the positive effect of making available this knowledge of word forms.

Even though practicing items sometimes have no effect on VWM (Olson and Jiang, 2004; Eng et al., 2005), research has found that performance in a trained working memory task usually increases with practice (Redick et al., 2015). This was also found in VWM (Gaspar et al., 2013; Kundu et al., 2013; Owens et al., 2013; Schwarb et al., 2015) and also with Chinese characters (Zimmer et al., 2012). However, it has not yet been shown that

providing word form knowledge outside of a working memory task also enhances performance. We wanted to demonstrate this in a brain imaging experiment.

The neural consequences of practice can be various (Kelly and Garavan, 2005). In a task with low demands, Heinzel and colleagues observed that after training, participants exhibited less neural activity than before training to achieve the same level of performance. The authors explained this effect by higher neural efficiency (Heinzel et al., 2014, 2016). The same result was observed in a previous study using Chinese characters (Zimmer et al., 2012). A similar effect was also expected in this experiment, specifically in two regions: the intraparietal sulcus (IPS) and the fusiform cortex (FFC).

The IPS is a core structure of working memory (Xu, 2007; Xu and Chun, 2009; Xu and Jeong, 2016). It contributes to individuation and identification of stimuli in working memory tasks and is involved in focusing attention on memorized stimuli. Activity in the IPS was found to increase with set size and level off at working memory capacity (Xu and Chun, 2006; Robitaille et al., 2010). Hence, with small set sizes, we expected less neural activity for trained than for untrained characters. With larger set sizes, neural activity for both sets of items should approach a common maximum.

Additionally, according to the sensory recruitment hypothesis (Serences, 2016), those structures that encode the stimuli during perception should represent the items in working memory. For characters, this should be the VWFA in the FFC, as was shown for Western languages (Dehaene et al., 2002; Liu et al., 2008; Dehaene and Cohen, 2011) and for Chinese (Xue et al., 2006; Xue and Poldrack, 2007). If trained characters are recognized with less effort than untrained ones, we should see a reduction in neural activity within the VWFA.

In order to test this hypothesis, German participants (novices) learned the orthography of 12 Chinese characters without any information on pronunciation or meaning. We then compared memory for trained and untrained characters in a change detection task.

## Participants

Twenty-four students from Saarland University took part in the experiment. Two were excluded from the analysis because of below-chance performances. None of the participants had pre-experimental experience with Chinese language, and no participant took additional time to learn the word forms or their meanings outside of the practice sessions.

## Design and Procedure

Participants were presented repeatedly with a set of 12 characters in 12 training sessions (3 per week), in which participants saw animated writing videos and actively copied the characters. In developing the training materials, particular attention was paid to the guidelines that writing videos should establish “a high quality representation of the visual-spatial structure of the character and its orthography” (Cao et al., 2013a,b, p. 1670), especially when combined with animation (Xu et al., 2013). In each writing animation of the training material, one character was drawn stroke by stroke. The training sessions consisted of stroke

copying, character drawing, a one-back task, and, from session 7, additionally, a written free recall. Details can be found in the **Supplementary Material**.

Working memory performance was tested in a change detection task with a set size of one to three in a 3T scanner. In the pre-test before training, all items were novel; in the post-test, half were trained and half untrained. The items that were trained and tested as novel material were counterbalanced across subjects. Presentation time was 1,000 ms. After a retention interval of 4,000 ms, one central test stimulus was presented. The displayed “items” were characters or random patches, i.e., squares of the same size as the characters, filled with spatially randomized pixels of the characters. In the MR analyses, the parameters per set size were estimated as contrasts between the character condition and the random patch condition to remove non-specific visual activity.

## Results

### Behavioral Measures

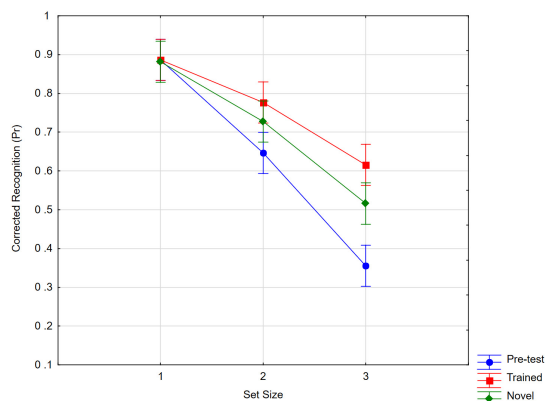
To test knowledge acquisition of the trained word forms, we analyzed participants' performance in the free recall task. In the first recall (session 7), participants could already write down from memory 10.6 of the 12 characters on average. By session 12, their performance had increased to 11.6 out of 12 [ $t(21) = 3.31, p < 0.005$ ].

As shown in **Figure 8**, Pr scores in the change detection task were analyzed in a  $3 \times 3$  analysis with set size (1, 2, 3) and status (pre-test, novel, trained) as the factors. The results showed a significant interaction [ $F(4,84) = 6.70, p < 0.001, \eta_p^2 = 0.24, MSE = 0.014, \epsilon = 0.85$ ]. Such an effect was also seen when we compared pre-test memory for novel characters with post-test memory [ $F(2,42) = 4.33, p < 0.05, \eta_p^2 = 0.17, MSE = 0.017, \epsilon = 0.85$ ]. This is an unexpected general effect of practice. Additionally, the results showed an item-specific training effect. Memory performance at set size one showed ceiling effects already in the pre-test, and we therefore only contrasted trained with novel characters at set sizes of two and three, trained characters were memorized better than novel ones [ $t(21) = 1.90, p < 0.05$ , single sided]. The estimated “capacity” in terms of the number of items that could be stored in working memory was 1.1 at pre-test and, at post-test, 1.5 for novel characters and 1.8 for trained characters.

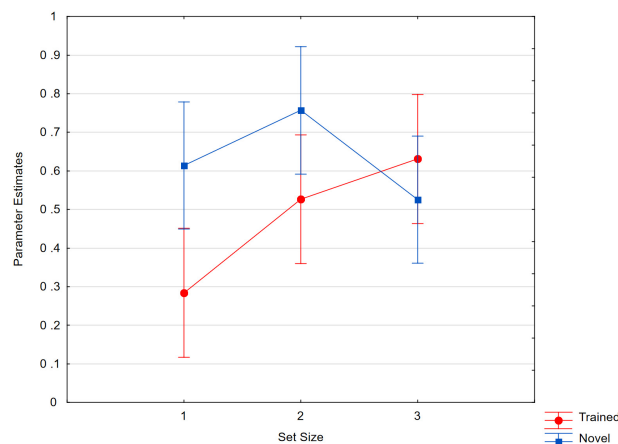
### Brain Imaging Data

Bold responses were analyzed in SPM 12. Pre-processing followed the standard procedure. In the design matrix of the model, only trials with correct responses were included. We were interested in set size and training effects during the retention interval within the IPS and FFC. For this analysis, we modeled the activity during the 4 s retention interval as a boxcar function using the canonical double-gamma hemodynamic response function (HRF) functions of SPM. Due to the predicted changes in specific brain areas, we estimated parameters for average activity within pre-specified regions of interest (ROIs) using the MarsBAR ROI toolbox. We extracted for each participant the parameter estimates for the six combinations of set size and item status within these ROIs and used these scores as dependent variables





**FIGURE 8 |** Recognition scores as a function of set size (1, 2, 3) and item status (pre-test, trained, novel). Whiskers show within subject confidence intervals based on the error term of the interaction according to Loftus and Masson (1994).



**FIGURE 10 |** Parameter estimates for activity in the sphere around the center of Cohen's VWF area as a function of set size and item status. Whiskers are within subject confidence intervals based on the error term of the interaction according to Loftus and Masson (1994).

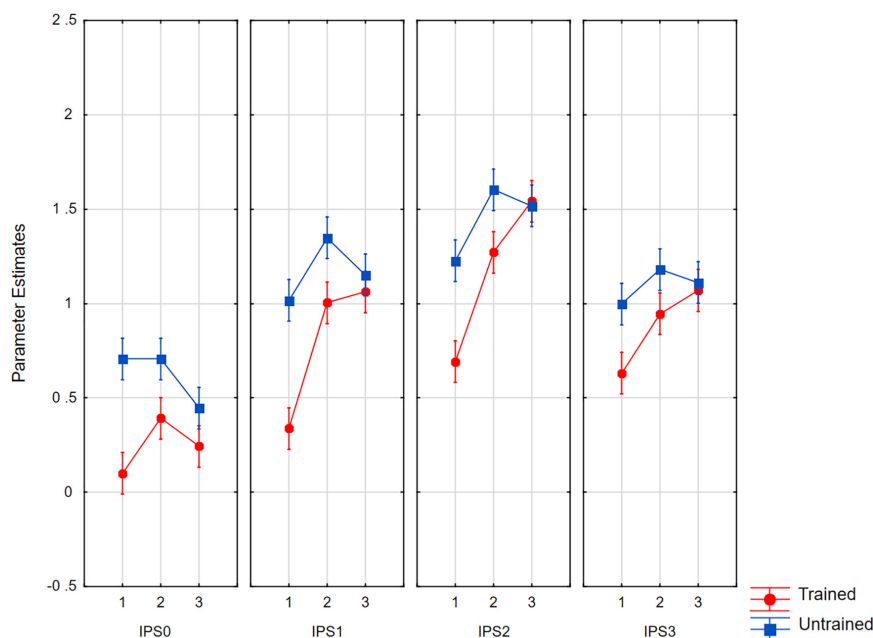
in the ANOVAs. For time course analysis within these ROIs, we extracted parameter estimates for nine repetition time (TR) points beginning with the retention interval. In the time course analysis, we modeled activity as finite impulse responses; this method does not constrain the shape of the time course.

### Task-Related Activity During Retention

Xu and Chun (2006) have highlighted the relevance of the inferior IPS (+26/−25, −65/−70, 34/29)<sup>1</sup> and superior IPS

(+26/−21, −52/−66, 45/46) for individuation and identification, respectively. Sheremata et al. (2010) subdivided the IPS into smaller regions, IPS0 to IPS4 and anterior IPS. With regard to visual short-term memory, they argued that set size drives activity in IPS0 (±26, −78, 29), IPS1 (±22, −71, 41), and IPS2 (±18, −63, 52). We therefore defined spheres with a diameter of 5 mm around the coordinates of IPS0 to IPS3 and extracted parameter estimates for average activity within these volumes for the different combinations of set size and character status. We

<sup>1</sup>Talairach coordinates in the left/right hemisphere.



**FIGURE 9 |** Parameter estimates for average activity in the left IPS0 to IPS3 volumes as a function of set size and item status. Whiskers show within subject confidence intervals based on the error term of the triple interaction according to Loftus and Masson (1994).

expected to see an increase in activity with set size but less activity for trained than for novel characters, while both would approach a common maximum at high set sizes. We only report activity in the left IPS: the right IPS showed qualitatively the same effects.

Averages per condition (**Figure 9**) were analyzed in a  $4 \times 2 \times 3$  ANOVA with the factors as follows: IPS region (0 to 3); training status (trained, novel); and set size (1, 2, 3). The results showed a significant effect of region [ $F(3,63) = 7.13$ ,  $p < 0.0005$ ,  $\eta_p^2 = 0.25$ ,  $MSE = 2.45$ ,  $\varepsilon = 0.81$ ], which interacted with set size [ $F(6,126) = 3.57$ ,  $p < 0.007$ ,  $\eta_p^2 = 0.15$ ,  $MSE = 0.247$ ,  $\varepsilon = 0.74$ ]. The interaction was due to the fact that the strongest set size effects were seen in IPS1 and IPS2, which roughly correspond to the inferior and superior IPS discussed by Xu and Chun (2006). With set sizes of one and two, activity was lower for trained than for novel material [ $F(1,21) = 9.41$ ,  $p < 0.006$ ]. For trained characters, the activity per set size followed the order  $1 < 2 < 3$ , whereas for untrained material, the order was  $1 < 2 = 3$ . In IPS0, only the training effect was significant [ $F(1,21) = 7.71$ ,  $p < 0.05$ ; set size:  $F < 1$ ].

The average activity in the FFC was estimated within a 5 mm sphere around the coordinates of the VWFA ( $-42$ ,  $-57$ ,  $-15$ ) as defined by Cohen et al. (2002). As shown in **Figure 10**, we analyzed the extracted values in a  $2 \times 3$  analysis with item status (trained, novel) and set size (1, 2, 3) as the factors. The only significant effect was found in the interaction between the two factors [ $F(2,42) = 4.33$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.17$ ,  $MSE = 0.132$ ,  $\varepsilon = 0.68$ ]. For set sizes of one and two (but not set size of three), trained characters showed less activity than novel ones [ $t(21) = 3.09$ ,  $p < 0.005$ ].

### Time Course of Activity

We then looked at the time course of neural activity within IPS1 and IPS2, the two IPS regions with the strongest effects. We analyzed the data in a  $2 \times 2 \times 3 \times 9$  ANOVA with the four factors as follows: region (IPS1, IPS2); item status (trained, novel); set size (1–3); and time point (T1 to T9). Because IPS1 and IPS2 showed comparable effects and no interactions, we present in **Figure 11** the mean parameter estimates for the average of both IPS regions in steps of TR (1,800 ms), time-locked to the onset of the retention interval. Trained characters elicited less activity in the IPS than novel characters [ $F(1,21) = 6.41$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.23$ ,  $MSE = 0.10$ ]. Activity increased with set size [ $F(2,42) = 7.71$ ,  $p < 0.005$ ,  $\eta_p^2 = 0.27$ ,  $MSE = 0.070$ ,  $\varepsilon = 0.94$ ] and decreased over time [ $F(8,168) = 28.23$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.57$ ,  $MSE = 0.069$ ,  $\varepsilon = 0.55$ ]. Set size interacted with time point [ $F(16,336) = 9.77$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.32$ ,  $MSE = 0.017$ ,  $\varepsilon = 0.57$ ], and the three-way interaction was also significant [ $F(16,336) = 2.03$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.09$ ,  $MSE = 0.003$ ,  $\varepsilon = 0.50$ ].

In **Figure 11**, neural activities at each time point are plotted separately for the three set sizes. With small set sizes, starting at the same maximum, the activity returned to baseline more quickly for trained characters than for novel characters. For a set size of three, no differences were observed.

For the analysis of time courses in the FFC, we defined the region of interest around the coordinates ( $-40$ ,  $-56$ ,  $-3$ ), which showed the maximum of activity in our participants. As shown in **Figure 12**, the data were analyzed in a  $2 \times 3 \times 9$  ANOVA

with item status (trained, novel), set size (1, 2, 3), and time point (1–9) as the factors. Trained items were associated with less activity than novel items [ $F(1,21) = 12.20$ ,  $p < 0.005$ ,  $\eta_p^2 = 0.37$ ,  $MSE = 0.008$ ]. Activity increased with set size [ $F(2,42) = 3.72$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.15$ ,  $MSE = 0.015$ ,  $\varepsilon = 0.90$ ] and decreased over time points [ $F(8,168) = 37.68$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.64$ ,  $MSE = 0.009$ ,  $\varepsilon = 0.44$ ]. Additionally, there were significant interactions between item status and time point [ $F(8,168) = 2.41$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.10$ ,  $MSE = 0.002$ ,  $\varepsilon = 0.55$ ] and between set size and time point [ $F(16,336) = 4.98$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.19$ ,  $MSE = 0.002$ ,  $\varepsilon = 0.51$ ]. The triple interactions were also significant [ $F(16,336) = 2.47$ ,  $p < 0.005$ ,  $\eta_p^2 = 0.11$ ,  $MSE = 0.002$ ,  $\varepsilon = 0.64$ ]. With regard to the FFC, at set sizes one and two, activity in trials with novel characters stayed high for longer than activity in trials with trained characters.

### Performance-Related Neural Activity

If neural activity is relevant to task performance, we should find a relationship with behavioral performance. Because the number of participants was too low for a correlational analysis, we compared neural activity of high and low performers separated at the median performance level in the trained character condition. High and low performers did not differ in performance at pre-test (0.61; 0.65), but at post-test, they differed for trained characters (0.87; 0.65) and for novel characters (0.79; 0.63). We then conducted our time course analyses with the additional factor of performance (high, low). We only analyzed the effects in IPS2 and the FFC, as they were the ROIs showing the strongest effects, and we do not report effects that were not moderated by the performance level factor.

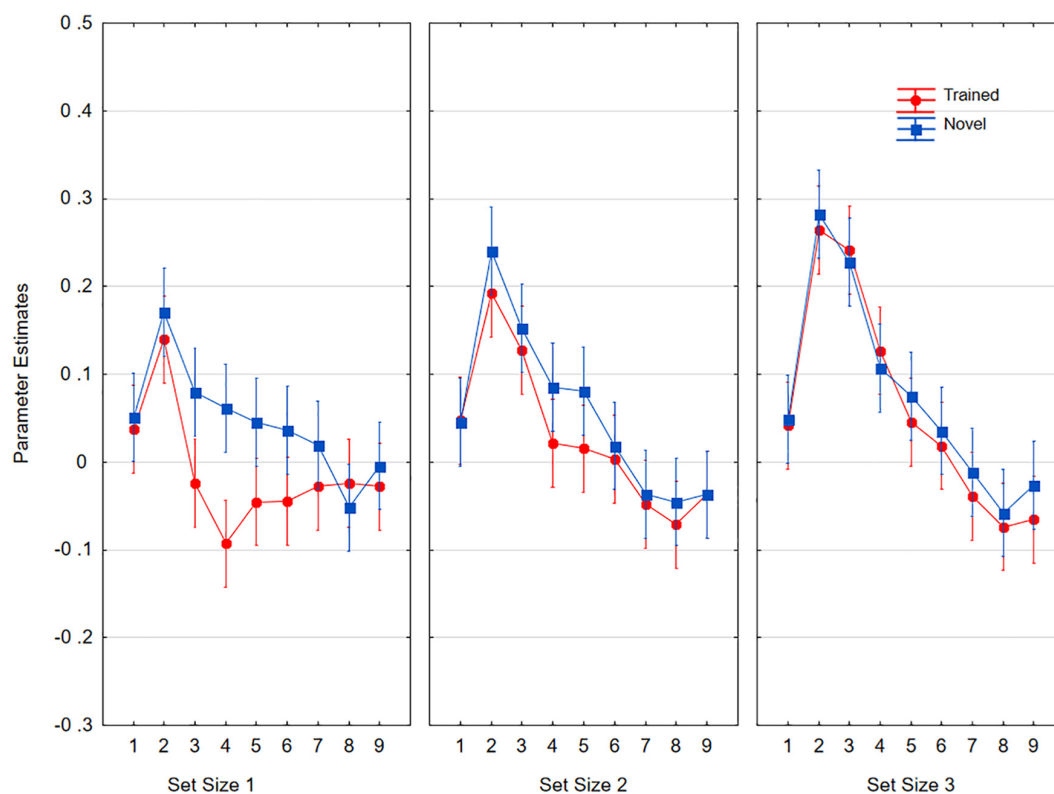
In the IPS, high-performing participants showed higher neural activity than low-performing participants [ $F(1,20) = 4.39$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.18$ ,  $MSE = 0.0175$ ]. This effect interacted with time point [ $F(8,160) = 3.32$ ,  $p < 0.005$ ,  $\eta_p^2 = 0.14$ ,  $MSE = 0.040$ ,  $\varepsilon = 0.52$ ]. The interaction with set size approached significance [ $F(2,40) = 2.88$ ,  $p < 0.06$ ,  $\eta_p^2 = 0.13$ ,  $MSE = 0.0396$ ]. The interaction is shown in **Figure 13**. Although we observed no higher-order interaction, **Figure 14** shows the highest resolution of data to illustrate the consistency of results across conditions.

In the fusiform ROI, the only interaction was found between time point and performance level [ $F(8,160) = 5.35$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.21$ ,  $MSE = 0.039$ ,  $\varepsilon = 0.59$ ], which was due to enhanced neural activity among high performers exclusively at time point T2 [ $F(1,20) = 19.97$ ,  $p < 0.001$ ]. No other effect with performance level approached significance.

### Discussion

The behavioral results of Experiment 5 showed a training effect only among the high performers. This was not an item-specific effect, because the performance difference was observed for trained as well as for novel characters. However, item-specific neural training effects were observed for all participants.

The IPS showed set size and training effects; additionally, high performers showed a noticeably stronger activity than low performers. We consider the activity in the IPS as being modulated by a participant's task engagement. The data suggest that, independent of the items' training status,



**FIGURE 11** | Time course of activity averaged across IPS1 and IPS 2 volumes as a function of item status (trained, novel) and time points (TR1 to TR9) for the three set sizes (1, 2, 3). Whiskers show within subject confidence intervals based on the error term of the triple interaction according to Loftus and Masson (1994).

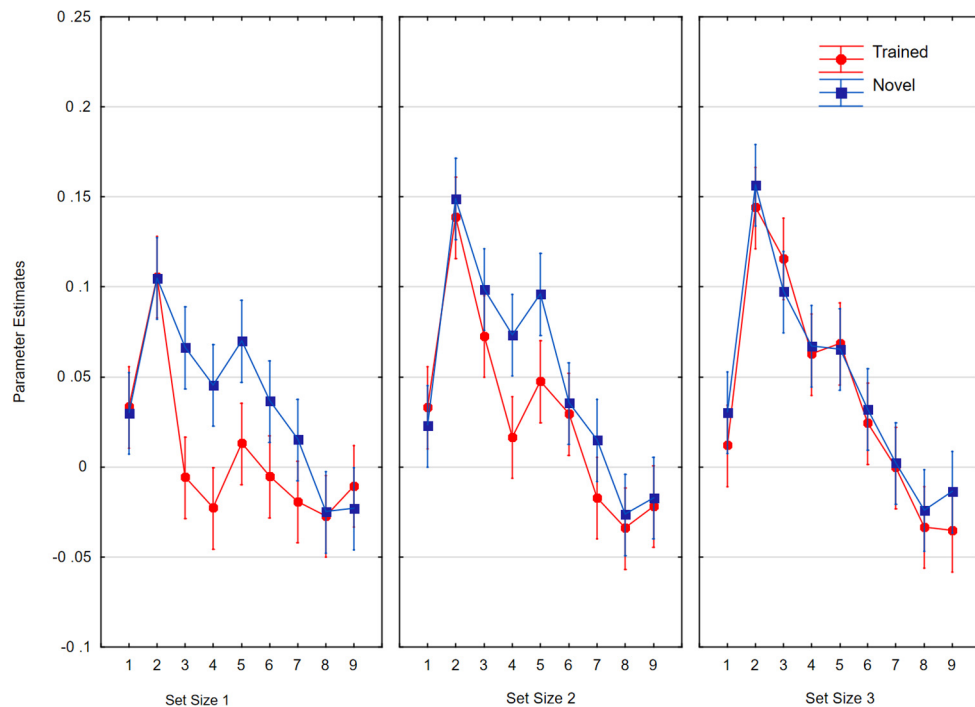
high performers put more effort into the task than low performers, which probably caused their general memory advantage. We can only speculate why low performers showed a neural but no behavioral training effect. It may be that they paid less attention to a study display if they recognized that it consisted of apparently easier trained items. Such a counterproductive adaptive behavior would use up training gains.

In the IPS and FFC, we saw similar time courses. We interpret the duration of elevated activity as a correlate of effort invested into character encoding and maintenance. In the beginning, in a stimulus-driven manner, set size had the main influence because word form information that provides the training effect only becomes available in the course of encoding. The total amount of effort that is invested into a study display should increase with set size up to the capacity limit. Familiar items needed less effort, which was observed as long as capacity was not exceeded. Although this is speculative, we interpret the fast decay of activity seen for trained characters as an indicator of a shortened or reduced encoding process and/or maintenance process. However, these are *post hoc* interpretations of the time courses, which also depend on the model used for analysis and experimental parameters. In further studies, we plan to vary the length of encoding time and/or of the retention interval in order to gain insight into the origins of time course changes. Additionally, to obtain an estimate of the cognitive effort

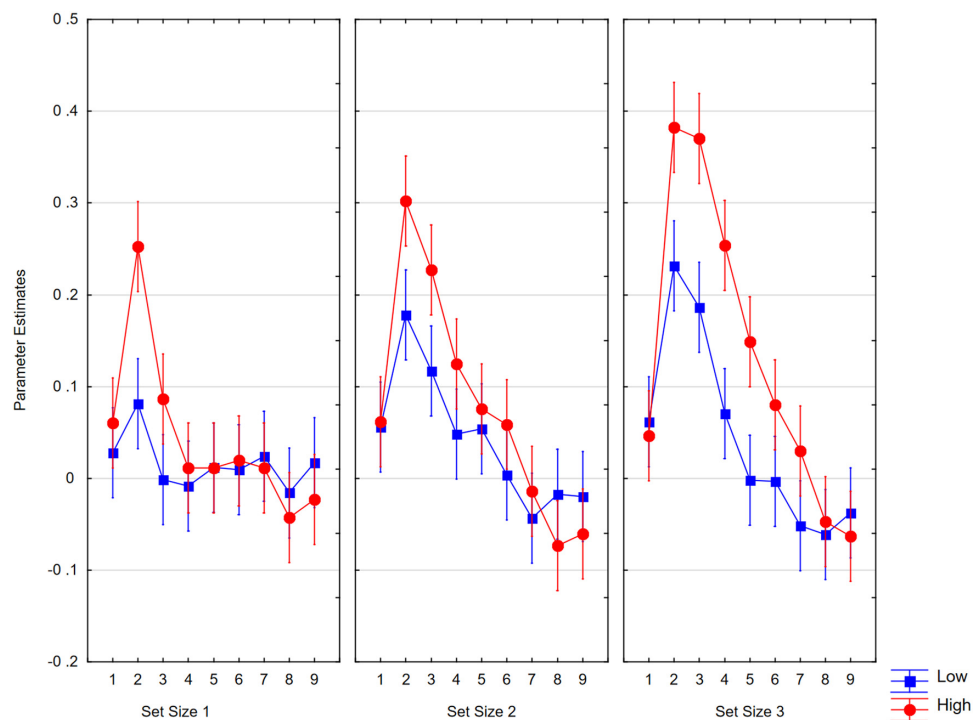
invested, we should add a independent measure of cognitive effort, for instance, pupil diameters (Kursawe and Zimmer, 2015; Unsworth and Robison, 2016).

## GENERAL DISCUSSION

The results of Experiments 1 to 3 demonstrate that Chinese literates had better working memory for Chinese characters than novices without word form knowledge. Experts could hold in memory 2.4 characters on average, whereas novices had a capacity of only one character. Similar results have been reported in other experiments (Sun et al., 2011; Ngiam et al., 2019). The strong effect of visual similarity together with the weak effect of semantic distractor similarity also suggests that experts hold characters as visual representations in VWM (Experiment 4). There was no difference between experts and novices in detecting changes to figures, whether the change was big or small (Experiments 1 and 2). In remembering specific details of characters, experts detected a character change much better than novices, but both groups did not differ in being poor at detecting changes to a character's font (Experiments 2 and 3) and in being poor at memorizing pseudo characters (Experiment 4). In sum, Chinese literates exhibited a better working memory performance for abstract Chinese word forms, but we did not find indicators of an enhanced working memory

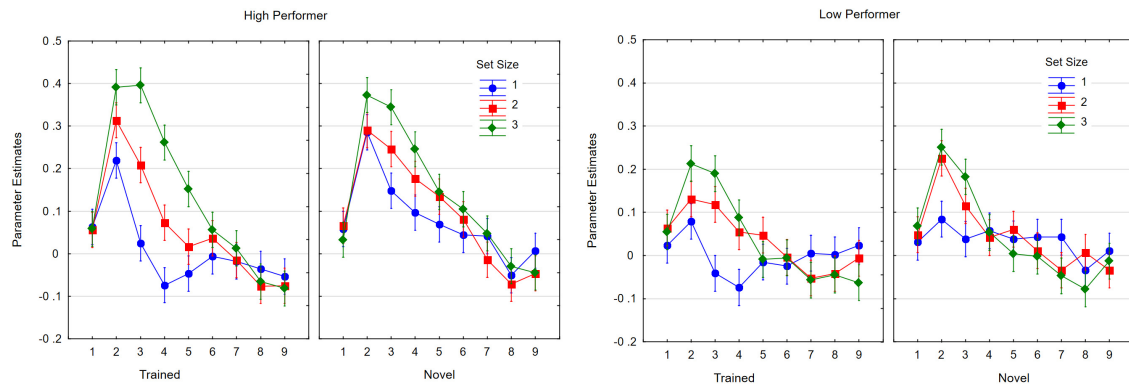


**FIGURE 12 |** Parameter estimates for the fusiform cluster as a function of item status (trained, novel) and time point (TR1 to TR9) for the three set sizes. Whiskers show within subjects confidence intervals based on the error term of the interaction according to Loftus and Masson (1994).



**FIGURE 13 |** Parameter estimates for activity in the IPS2 volume as a function of performance level (low, high) and time point (TR1 to TR9) for the three set sizes. Whiskers show within subject confidence intervals based on the error term of the interaction according to Loftus and Masson (1994).





**FIGURE 14 |** Parameter estimates for activity in the IPS2 volume as a function of time point (TR1 to TR9) and set size (1, 2, 3) when high performers (**left**) or low performers (**right**) memorized trained or novel characters. Whiskers show within subject confidence intervals based on the error term of the highest interaction according to Loftus and Masson (1994).

in terms of the number of remembered objects and visual features or visual details.

We assume that Chinese literates represent characters as tokens in working memory, each making reference to a word form and additionally representing some coarse episodic information, e.g., color (Gao et al., 2013). Word forms make generic orthographic details available as templates (Dehaene et al., 2005) and allow literates to identify characters quickly. Consequentially, for familiar alphabets, it is not necessary to encode all perceptual details. The important variable for character encoding in working memory is therefore participants' familiarity with an alphabet (Ngiam et al., 2019). The memory demand of a task is a function of the characters' perceived complexity, i.e., familiarity (Dall and Sorensen, 2019). Furthermore, a pattern that is partially forgotten can be completed from long-term memory (Thalmann et al., 2019). Comparison errors at the test stage would be unlikely if word forms assign the remembered and perceived character to different categories (words). A possibility to test this would be confidence ratings (Xie and Zhang, 2017a), which could provide evidence that experts' memory is associated with recollection, whereas novices rely on familiarity. Because word forms do not represent low-level information, experts have no advantage for font memory; for a similar argument regarding visual search, see Baier and Ansorge (2019). Also, pseudo characters with no long-term representation have no expertise advantage. It seems that experts can reliably memorize only one complex unknown character, which is no different from novices. Olsson and Poom (2005) make a similar suggestion concerning another type of material with no categories in long-term memory.

Novices do not have word form knowledge available, and this might render the available encoding time a moderator for the size of the expertise advantage in working memory. This can induce a disadvantage exclusively for novices if they have insufficient time to represent the items in working memory. In a series of experiments using familiar and unfamiliar Pokémons as stimuli, Xie and colleagues reported results that favor this possibility. They observed higher memory capacity for familiar

than for unfamiliar figures if the encoding time was constrained (Xie and Zhang, 2017b). At 117 ms encoding time, both item types registered equally poor memory, but with a study time of 314 or 500 ms, familiar Pokémons were remembered better than unfamiliar ones. When the time was increased further to 1,000 ms, the difference disappeared (Xie and Zhang, 2017c, 2018). This suggests that the encoding time provided is the limiting factor. However, other studies could not replicate this. Ngiam et al. (2019) investigated memory for different alphabets. From 120 to 270 ms, performance for all fonts increased. Familiar fonts (Courier, Helvetica, Bookman) had an encoding rate of about 42 items per second and a capacity of about 4 items. Unfamiliar fonts (the handwriting style of the font Künstler; Braille; Hebrew; and Chinese characters) had lower encoding rates (7–15) and capacities (1–1.8). However, the effects of familiarity remained constant between 270 and 600 ms. In a series of experiments, Sun et al. (2011) varied encoding time between 217 and 683 ms. Performance increased up to 450 ms, and the positive effect of familiarity was independent of encoding time. In the experiments reported in this paper, familiar items also had a memory advantage independent of encoding time (500 or 1,000 ms). To conclude, with encoding times of 500 ms and above, working memory for characters seems no longer to depend on encoding time but, rather, on item familiarity. The cancelation of the familiarity effect for Pokémons under long study times may be specific to such material. Visual inspection of the Pokémons suggests that they are complex but perceptually more different than unfamiliar letters. Hence, with long encoding time, novices may also be able to represent unique features of Pokémons but not of characters that allow discrimination between items in working memory.

Another critical issue is the code used in working memory. We have claimed that we always observe effects of visual working memory, although Chinese literates have available phonological and semantic information on the characters. No doubt, people can recode items and use different codes (e.g., Lewis-Peacock et al., 2015). However, in a standard visual change detection task with hundreds of trials and short trial lengths, recoding seems not

to be a strategy. The presence of articulatory suppression or of a verbal preload had no effect or only a marginal influence on the results of several experiments (Luck and Vogel, 1997; Morey and Cowan, 2004; Pelli et al., 2006; Sun et al., 2011; Xie and Zhang, 2017c). We observed that visual but not semantic similarity of foils impaired memory (Experiment 4). All of these results suggest a visual code. When verbal rehearsal was a likely strategy (Yu et al., 1985; Zhang and Simon, 1985), other results were observed. For example, if Chinese literates saw items sequentially presented at a slow pace (750 ms per item) and their memory was tested by a written recall, memory performance increased to seven characters, much higher than the “capacity” of VWM. If phonological recoding was rendered useless—the items were homophones or radicals without pronunciations—performance dropped to about three items, and the authors interpreted the smaller set of items as reflecting visual short-term memory. Hence, in change detection tasks, even literates in Chinese should represent characters as visual information in working memory.

According to the sensory recruitment hypothesis of working memory, the visual representations should be represented in the same neural structures that encode the information. Therefore, for characters, activity should be seen in the FFC and the IPS, as these are involved in encoding and attentional control, respectively. The intensity of the neural signal should be a function of the necessary encoding and maintenance effort. In Experiment 5, therefore, trained material showed less neural activity than novel characters in the FFC and IPS. Activity increased with set size, approached a common maximum that was independent of training status, and lasted longer for higher set sizes. Whether the changed time courses are effects of encoding or maintenance or both cannot be decided. To test this, encoding and maintenance effort must be varied experimentally. Another result was that high performers generally showed more neural activity than low performers in the IPS, which we considered as task engagement. This conclusion also has to be substantiated by further empirical data.

In closing, some limitations of the brain imaging study need to be mentioned. We only compared neural activity in two regions for which we had specific hypotheses. However, working memory is provided by a network (Zimmer, 2008), which involves many regions that were not analyzed. Furthermore, we opted for a univariate approach in looking at the size of neural activity, not at its distribution (Postle, 2015). It is very likely that neural representations contribute to memory but are not expressed in elevated activity and are therefore not visible in a univariate approach. In a similar way, phonological or semantic information can also contribute to memory as part of a distributed network by making item representations distinct without having an influence that can be confirmed directly in elevated activity of an averaged signal or an overt behavioral measure. The phonology and meaning associated to a word form, for example, may stabilize a representation (Perfetti et al., 2013), even if these codes are not used to represent and address items in working memory. This might explain why word form training enhanced memory but did not allow the same memory performance as long experience with Chinese characters. Pelli et al. (2006) observed that after about 3,000 trials, Chinese illiterates reached the same efficiency

in letter identification as experts but not the same memory performance. It was still one versus three items. In Experiment 5, after word form training, German participants remembered 1 novel and about 1.8 trained characters on average, whereas Chinese native speakers held 2.4 items in VWM. These results suggest that long experience causes a reorganization of character representations which enhances memory. Probably, it makes item representations more distinct and, in this way, reduces interference in working memory.

However, independent of these effects of prolonged practice, our experiments have shown that perceptual long-term knowledge makes an important contribution to the expertise advantage in VWM. We assume that experts can quickly assign perceived items within their field of expertise to a known category, allowing them unique representations in working memory. These categories assign even highly similar items perceptually to different tokens if they differ in at least one critical feature. These categories allow them to reject changed items if the change alters the category. In other words, due to the category change, even a perceptually small change becomes a big change that can be detected easily. We suggest that this mechanism provides the good memory performance and the seemingly high resolution that we also see in visual working memory experiments with experts in other fields of expertise.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics committee of Fakultät für Empirische Humanwissenschaften. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

HZ and BF designed the experiments and conducted all statistical analyses. BF collected the data of Experiment 2 and 5 and analyzed the brain imaging data. HZ mainly wrote the manuscript. Both authors edited the final draft and agreed on its content.

## FUNDING

This research was supported by a grant from the Deutsche Forschungsgemeinschaft to the International Research Training Group Adaptive Minds (IRTG 1457) at Saarland University, Saarbrücken, Germany, and the Institute of Psychology at the Chinese Academy of Sciences, Beijing, China. We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and Saarland University within the funding program Open Access Publishing.

## ACKNOWLEDGMENTS

We would like to thank Xiaolan Fu for their advice in selecting the Chinese language material and Zhao Ke for running the experiments in China.

## REFERENCES

- Alvarez, G. A., and Cavanagh, P. (2004). The capacity of visual short term memory is set both by visual information load and by number of objects. *Psychol. Sci.* 15, 106–111. doi: 10.1111/j.0963-7214.2004.01502006.x
- Alvarez, G. A., and Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Percept. Psychophys.* 70, 346–364. doi: 10.3758/pp.70.2.346
- Awh, E., Barton, B., and Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychol. Sci.* 18, 622–628. doi: 10.1111/j.1467-9280.2007.01949.x
- Baier, D., and Ansorge, U. (2019). Investigating the role of verbal templates in contingent capture by color. *Atten. Percept. Psychophys.* 81, 1846–1879. doi: 10.3758/s13414-019-01701-y
- Barton, B., Ester, E. F., and Awh, E. (2009). Discrete resource allocation in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1359–1367. doi: 10.1037/a0015792
- Bourassa, D. C., and Besner, D. (1994). Beyond the articulatory loop: a semantic contribution to serial order recall of subspan lists. *Psychon. Bull. Rev.* 1, 122–125. doi: 10.3758/bf03200768
- Brady, T. F., Konkle, T., Oliva, A., and Alvarez, G. A. (2009). Detecting changes in real-world objects: the relationship between visual long-term memory and change blindness. *Commun. Integr. Biol.* 2, 1–3. doi: 10.4161/cib.2.1.7297
- Brady, T. F., Störmer, V. S., and Alvarez, G. A. (2016). Working memory is not fixed-capacity: more active storage capacity for real-world objects than for simple stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7459–7464. doi: 10.1073/pnas.1520027113
- Brady, T. F., Störmer, V. S., Shafer-Skelton, A., Williams, J. R., Chapman, A. F., and Schill, H. (2019). Scaling up visual attention and visual working memory to the real world. *Psychol. Learn. Motiv.* 70, 29–69. doi: 10.1016/bs.plm.2019.03.001
- Cao, F., Rickles, B., Vu, M., Zhu, Z., Chan, D. H. L., Harris, L. N., et al. (2013a). Early stage visual-orthographic processes predict long-term retention of word form and meaning: a visual encoding training study. *J. Neurolinguistics* 26, 440–461. doi: 10.1016/j.jneuroling.2013.01.003
- Cao, F., Vu, M., Chan, D. H. L., Lawrence, J. M., Harris, L. N., Guan, Q., et al. (2013b). Writing affects the brain network of reading in Chinese: a functional magnetic resonance imaging study. *Hum. Brain Mapp.* 34, 1670–1684. doi: 10.1002/hbm.22017
- Chen, H.-C., Flores d'Arcais, G. B., and Cheung, S.-L. (1995). Orthographic and phonological activation in recognizing Chinese characters. *Psychol. Res.* 58, 144–153. doi: 10.1007/bf00571102
- Chen, Y. C., and Yeh, S. L. (2015). Binding radicals in Chinese character recognition: evidence from repetition blindness. *J. Mem. Lang.* 78, 47–63. doi: 10.1016/j.jml.2014.10.002
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., and Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* 125, 1054–1069. doi: 10.1093/brain/awf094
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–185. doi: 10.1017/s0140525x01003922
- Curby, K. M., and Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychon. Bull. Rev.* 14, 620–628. doi: 10.3758/bf03196811
- Curby, K. M., Glazek, K., and Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 94–107. doi: 10.1037/0096-1523.35.1.94
- Dall, J. O., and Sorensen, T. A. (2019). “Short-term memory capacity is dependent on perceived complexity, and not visual complexity per se,” in *Proceedings of Annual OAK Meeting*, Aarhus.
- Dehaene, S., and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends Cogn. Sci.* 15, 254–262. doi: 10.1016/j.tics.2011.04.003
- Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. *Trends Cogn. Sci.* 9, 335–341. doi: 10.1016/j.tics.2005.05.004
- Dehaene, S., Le Clec'h, G., Poline, J. B., Le Bihan, D., and Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport* 13, 321–325. doi: 10.1097/00001756-200203040-00015
- Demetriou, A., Kui, Z. X., Spanoudis, G., Christou, C., Kyriakides, L., and Platsidou, M. (2005). The architecture, dynamics, and development of mental processing: Greek, Chinese, or Universal? *Intelligence* 33, 109–141. doi: 10.1016/j.intell.2004.10.003
- Eng, H. Y., Chen, D., and Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychon. Bull. Rev.* 12, 1127–1133. doi: 10.3758/bf03206454
- Fukuda, K., Vogel, E., Mayr, U., and Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychon. Bull. Rev.* 17, 673–679. doi: 10.3758/17.5.673
- Gao, Z., Ding, X., Yang, T., Liang, J., and Shui, R. (2013). Coarse-to-fine construction for high-resolution representation in visual working memory. *PLoS One* 8:e57913. doi: 10.1371/journal.pone.0057913
- Gao, Z., Li, J., Liang, J., Chen, H., Yin, J., and Shen, M. (2009). Storing fine detailed information in visual working memory—evidence from event-related potentials. *J. Vis.* 9:17. doi: 10.1167/9.7.17
- Gao, Z., Li, J., Yin, J., and Shen, M. (2010). Dissociated mechanisms of extracting perceptual information into visual working memory. *PLoS One* 5:e14273. doi: 10.1371/journal.pone.0014273
- Gaspar, J. G., Neider, M. B., Simons, D. J., McCarley, J. S., and Kramer, A. F. (2013). Change detection: training and transfer. *PLoS One* 8:e67781. doi: 10.1371/journal.pone.0067781
- Heinzel, S., Lorenz, R. C., Brockhaus, W. R., Wüstenberg, T., Kathmann, N., Heinz, A., et al. (2014). Working memory load-dependent brain response predicts behavioral training gains in older adults. *J. Neurosci.* 34, 1224–1233. doi: 10.1523/JNEUROSCI.2463-13.2014
- Heinzel, S., Lorenz, R. C., Pelz, P., Heinz, A., Walter, H., Kathmann, N., et al. (2016). Neural correlates of training and transfer effects in working memory in older adults. *Neuroimage* 134, 236–249. doi: 10.1016/j.neuroimage.2016.03.068
- Hue, C. W., and Erickson, J. R. (1988). Short-term memory for Chinese characters and radicals. *Mem. Cognit.* 16, 196–205. doi: 10.3758/bf03197752
- Jackson, M. C., and Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 556–568. doi: 10.1037/0096-1523.34.3.556
- Kelly, A. M., and Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cereb. Cortex* 15, 1089–1102. doi: 10.1093/cercor/bhi005
- Kundu, B., Sutterer, D. W., Emrich, S. M., and Postle, B. R. (2013). Strengthened effective connectivity underlies transfer of working memory training to tests of short-term memory and attention. *J. Neurosci.* 33, 8705–8715. doi: 10.1523/JNEUROSCI.5565-12.2013
- Kursawe, M. A., and Zimmer, H. D. (2015). Costs of storing colour and complex shape in visual working memory: insights from pupil size and slow waves. *Acta Psychol.* 158, 67–77. doi: 10.1016/j.actpsy.2015.04.004
- Leck, K. J., Weekes, B. S., and Chen, M. J. (1995). Visual and phonological pathways to the lexicon: evidence from Chinese readers. *Mem. Cognit.* 23, 468–476. doi: 10.3758/BF03197248
- Lewis-Peacock, J. A., Drysdale, A. T., and Postle, B. R. (2015). Neural evidence for the flexible control of mental representations. *Cereb. Cortex* 25, 3303–3313. doi: 10.1093/cercor/bhu130
- Liu, C., Zhang, W. T., Tang, Y. Y., Mai, X. Q., Chen, H. C., Tardif, T., et al. (2008). The Visual Word Form Area: evidence from an fMRI study of implicit

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00516/full#supplementary-material>

- processing of Chinese characters. *Neuroimage* 40, 1350–1361. doi: 10.1016/j.neuroimage.2007.10.014
- Loftus, G. R., and Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychon. Bull. Rev.* 1, 476–490. doi: 10.3758/BF03210951
- Lorenz, E. S., Pratte, M. S., Angeloni, C. F., and Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Atten. Percept. Psychophys.* 76, 1975–1984. doi: 10.3758/s13414-014-0653-z
- Luck, S. J. (2008). “Visual short-term memory,” in *Visual Memory*, eds S. J. Luck, and A. Hollingworth, (Oxford: Oxford University Press), 43–85.
- Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846
- Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006
- Ma, M. Y., and Chuang, H. C. (2015). How form and structure of Chinese characters affect eye movement control. *J. Eye Mov. Res.* 8,
- Mance, I., and Vogel, E. K. (2013). Visual working memory. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 179–190. doi: 10.1002/wcs.1219
- Mate, J., Allen, R. J., and Baqués, J. (2012). What you say matters: exploring visual-verbal interactions in visual working memory. *Q. J. Exp. Psychol.* 65, 395–400. doi: 10.1080/17470218.2011.644798
- McBride, C., and Wang, Y. (2015). Learning to read Chinese: universal and unique cognitive cores. *Child Dev. Perspect.* 9, 196–200. doi: 10.1111/cdep.12132
- McBride-Chang, C., Chow, B. W. Y., Zhong, Y., Burgess, S., and Hayward, W. G. (2005). Chinese character acquisition and visual skills in two Chinese scripts. *Read. Writ.* 18, 99–128. doi: 10.1007/s11145-004-7343-5
- McBride-Chang, C., Zhou, Y., Cho, J.-R., Aram, D., Levin, I., and Tolchinsky, L. (2011). Visual spatial skill: a consequence of learning to read? *J. Exp. Child Psychol.* 109, 256–262. doi: 10.1016/j.jecp.2010.12.003
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology* 4, 61–64. doi: 10.20982/tqmp.04.2.p061
- Morey, C. C., and Cowan, N. (2004). When visual and verbal memories compete: evidence of cross-domain limits in working memory. *Psychon. Bull. Rev.* 11, 296–301. doi: 10.3758/bf03196573
- Morey, R. D., Morey, C. C., Brisson, B., and Tremblay, S. (2012). A critical evaluation of *c* as a measure of mnemonic resolution. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 1069–1072. doi: 10.1037/a0027335
- Ngiam, W. X. Q., Khaw, K. L. C., Holcombe, A. O., and Goodbourn, P. T. (2019). Visual working memory for letters varies with familiarity but not complexity. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 1761–1775. doi: 10.1037/xlm0000682
- Oberauer, K., and Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Mem. Cognit.* 41, 1212–1227. doi: 10.3758/s13421-013-0333-6
- Olson, I. R., and Jiang, Y. (2004). Visual short-term memory is not improved by training. *Mem. Cognit.* 32, 1326–1332. doi: 10.3758/bf03206323
- Olsson, H., and Poom, L. (2005). Visual memory needs categories. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8776–8780. doi: 10.1073/pnas.0500810102
- Owens, M., Koster, E. H. W., and Derakshan, N. (2013). Improving attention control in dysphoria through cognitive training: transfer effects on working memory capacity and filtering efficiency. *Psychophysiology* 50, 297–307. doi: 10.1111/psyp.12010
- Pelli, D. G., Burns, C. W., Farell, B., and Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Res.* 46, 4646–4674. doi: 10.1016/j.visres.2006.04.023
- Perfetti, C., Cao, F., and Booth, J. (2013). Specialization and universals in the development of reading skill: how Chinese research informs a universal science of reading. *Sci. Stud. Read.* 17, 5–21. doi: 10.1080/10888438.2012.689786
- Perfetti, C. A., Liu, Y., and Tan, L. H. (2005). The lexical constituency model: some implications of research on Chinese for general theories of reading. *Psychol. Rev.* 112, 43–59. doi: 10.1037/0033-295X.112.1.43
- Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Curr. Opin. Behav. Sci.* 1, 40–46. doi: 10.1016/j.cobeha.2014.08.004
- Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., and Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educ. Psychol. Rev.* 27, 617–633. doi: 10.1007/s10648-015-9314-6
- Robitaille, N., Marois, R., Todd, J., Grimault, S., Cheyne, D., and Jolicoeur, P. (2010). Distinguishing between lateralized and nonlateralized brain activity associated with visual short-term memory: fMRI, MEG, and EEG evidence from the same observers. *Neuroimage* 53, 1334–1345. doi: 10.1016/j.neuroimage.2010.07.027
- Schurkin, M. W., and Brady, T. F. (2019). When “capacity” changes with set size: ensemble representations support the detection of across-category changes in visual working memory. *J. Vis.* 19:3. doi: 10.1167/19.5.3
- Schwarb, H., Nail, J., and Schumacher, E. H. (2015). Working memory training improves visual short-term memory capacity. *Psychol. Res.* 80, 128–148. doi: 10.1007/s00426-015-0648-y
- Scolari, M., Vogel, E. K., and Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychon. Bull. Rev.* 15, 215–222. doi: 10.3758/pbr.15.1.215
- Sense, F., Morey, C. C., Prince, M., Heathcote, A., and Morey, R. D. (2017). Opportunity for verbalization does not improve visual change detection performance: a state-trace analysis. *Behav. Res. Methods* 49, 853–862. doi: 10.3758/s13428-016-0741-1
- Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Res.* 128, 53–67. doi: 10.1016/j.visres.2016.09.010
- Sheremata, S. L., Bettencourt, K. C., and Somers, D. C. (2010). Hemispheric asymmetry in visuotopic posterior parietal cortex emerges with visual short-term memory load. *J. Neurosci.* 30, 12581–12588. doi: 10.1523/JNEUROSCI.2689-10.2010
- Shivde, G., and Anderson, M. C. (2011). On the existence of semantic working memory: evidence for direct semantic maintenance. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1342–1370. doi: 10.1037/a0024832
- Simon, H. A. (1974). How big is a chunk? *Science* 183, 482–488. doi: 10.1126/science.183.4124.482
- Sorensen, T. A., and Kyllingsbaek, S. (2012). Short-term storage capacity for visual objects depends on expertise. *Acta Psychol.* 140, 158–163. doi: 10.1016/j.actpsy.2012.04.004
- Sun, H., Zimmer, H. D., and Fu, X. (2011). The influence of expertise and of physical complexity on visual short-term memory consolidation. *Q. J. Exp. Psychol.* 64, 707–729. doi: 10.1080/17470218.2010.511238
- Sze, W. P., Yap, M. J., and Rickard Liow, S. J. (2015). The role of lexical variables in the visual recognition of Chinese characters: a megastudy analysis. *Q. J. Exp. Psychol.* 68, 1541–1570. doi: 10.1080/17470218.2014.985234
- Thalmann, M., Souza, A. S., and Oberauer, K. (2019). How does chunking help working memory? *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 37–55. doi: 10.1037/xlm0000578
- Unsworth, N., and Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cogn. Affect. Behav. Neurosci.* 16, 601–615. doi: 10.3758/s13415-016-0417-4
- Vogel, E. K., Woodman, G. F., and Luck, J. (2001). Storage of features, conjunctions, and objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 92–114. doi: 10.1037/0096-1523.27.1.92
- Wagar, B. M., and Dixon, M. J. (2005). Past experience influences object representation in working memory. *Brain Cogn.* 57, 248–256. doi: 10.1016/j.bandc.2004.08.054
- Xie, W., and Zhang, W. (2017a). Dissociations of the number and precision of visual short-term memory representations in change detection. *Mem. Cognit.* 45, 1423–1437. doi: 10.3758/s13421-017-0739-7
- Xie, W., and Zhang, W. (2017b). Familiarity increases the number of remembered Pokemon in visual short-term memory. *Mem. Cognit.* 45, 677–689. doi: 10.3758/s13421-016-0679-7
- Xie, W., and Zhang, W. (2017c). Familiarity speeds up visual short-term memory consolidation. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 1207–1221. doi: 10.1037/xhp0000355
- Xie, W., and Zhang, W. (2018). Familiarity speeds up visual short-term memory consolidation: electrophysiological evidence from contralateral delay activities. *J. Cogn. Neurosci.* 30, 1–13. doi: 10.1162/jocn\_a\_01188
- Xu, Y. (2007). The role of the superior intraparietal sulcus in supporting visual short-term memory for multifeature objects. *J. Neurosci. Methods* 27, 1167–1168. doi: 10.1523/JNEUROSCI.3545-07.2007
- Xu, Y., Chang, L.-Y., Zhang, J., and Perfetti, C. A. (2013). Reading, writing, and animation in character learning in Chinese as a foreign language. *Foreign Lang. Ann.* 46, 423–444. doi: 10.1111/flan.12040



- Xu, Y., and Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440, 91–95. doi: 10.1038/nature04262
- Xu, Y., and Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends Cogn. Sci.* 13, 167–174. doi: 10.1016/j.tics.2009.01.008
- Xu, Y., and Jeong, S. K. (2016). “The contribution of human superior intraparietal sulcus to visual short-term memory and perception,” in *Mechanisms of Sensory Working Memory*, eds P. Jolicoeur, C. Lefebvre, and J. Martinez-Trujillo, (San Diego, CA: Academic Press), 33–42. doi: 10.1016/b978-0-12-801371-7.00004-1
- Xue, G., Chen, C., Jin, Z., and Dong, Q. (2006). Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. *Neuroimage* 31, 1315–1326. doi: 10.1016/j.neuroimage.2005.11.055
- Xue, G., and Poldrack, R. A. (2007). The neural substrates of visual perceptual learning of words: implications for the visual word form area hypothesis. *J. Cogn. Neurosci.* 19, 1643–1655. doi: 10.1162/jocn.2007.19.10.1643
- Yeh, S. L., Li, J. L., Takeuchi, T., Sun, V. C., and Liu, W. R. (2003). The role of learning experience on the perceptual organization of Chinese characters. *Vis. Cogn.* 10, 729–764. doi: 10.1080/13506280344000077
- Yu, B., Zhang, W., Jing, Q., Peng, R., Zhang, G., and Simon, H. A. (1985). STM capacity for Chinese and English language materials. *Mem. Cognit.* 13, 202–207. doi: 10.3758/bf03197682
- Yu, L., Zhang, Q., Priest, C., Reichle, E. D., and Sheridan, H. (2018). Character-complexity effects in Chinese reading and visual search: a comparison and theoretical implications. *Q. J. Exp. Psychol.* 71, 140–151. doi: 10.1080/17470218.2016.1272616
- Zhang, G., and Simon, H. A. (1985). STM capacity for Chinese words and idioms: chunking and acoustical loop hypotheses. *Mem. Cognit.* 13, 193–201. doi: 10.3758/bf03197681
- Zimmer, H. D. (2008). Visual and spatial working memory: from boxes to networks. *Neurosci. Biobehav. Rev.* 32, 1373–1395. doi: 10.1016/j.neubiorev.2008.05.016
- Zimmer, H. D., and Fu, X. (2008). “Working memory capacity and culture-based expertise,” in *Proceedings of the XXIV International Congress of Psychology*, Berlin.
- Zimmer, H. D., Popp, C., Reith, W., and Krick, C. (2012). Gains of item-specific training in visual working memory and their neural correlates. *Brain Res.* 1466, 44–55. doi: 10.1016/j.brainres.2012.05.019

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zimmer and Fischer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

@frontiersin



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership