The Effects of Similarity on High-Level Visual Working Memory Processing

Li Yang [1, 2, 3] *and Lei Mo* [1, 2, 3]

[1] Centre for Studies of Psychological Application, South China Normal University, Guangzhou, China

[2] School of Psychology, South China Normal University, Guangzhou, P. R. China

[3] Guangdong Key Laboratory of Mental Health and Cognitive Science,
  South China Normal University, Guangzhou, P. R. China

ABSTRACT

Similarity has been observed to have opposite effects on visual working memory (VWM) for complex images. How can these discrepant results be reconciled? To answer this question, we used a change-detection paradigm to test visual working memory performance for multiple real-world objects. We found that working memory for moderate similarity items was worse than that for either high or low similarity items. This pattern was unaffected by manipulations of stimulus type (faces vs. scenes), encoding duration (limited vs. self-paced), and presentation format (simultaneous vs. sequential). We also found that the similarity effects differed in strength in different categories (scenes vs. faces). These results suggest that complex real-world objects are represented using a *centre-surround inhibition organization*. These results support the *category-specific cortical resource* theory and further suggest that centre-surround inhibition organization may differ by category.

THE EFFECTS OF INTER-ITEM SIMILARITY ON VISUAL WORKING MEMORY FOR COMPLEX IMAGES

Appropriate processing of visual similarity is of vital importance to our representation of the world. Visual similarity affects multiple cognitive functions, such as object recognition, generalization, and categorization (Mate & Baqués, 2009). Over the past decade, similarity effects have been well established in the field of visual attention. Attention models incorporating *centre-surround inhibition* have been developed to account for these similarity effects.

In the domain of visual location-based attention, accuracy in a recognition task is enhanced when location similarity is high, decreases when location similarity reaches intermediate levels, and finally recovers when location similarity is low (Cutzu & Tsotsos, 2003; Hopf et al., 2006; Müller, Mollenhauer, Rösler, & Kleinschmidt, 2005). A similar pattern is observed in the domain of visual feature-based attention (Störmer & Alvarez, 2014). This finding suggests a selection profile: An excitatory peak is surrounded by a narrow inhibitory zone to limit interference in a feature space. Taken together, these findings reveal a centre-surround inhibition organization for visual attention.

The same centre-surround selection mechanism also maintains internally activated representations in visual working memory (VWM). Kiyonaga and Egner's (2016) research showed centre-surround inhibition organization in VWM for the first time. They sequentially presented two circles that varied in similarity to each other in circular colour space. They asked the participants to respond to whether the probed circle was the same as the cued one. They found that the recognition response time followed an inverted u-shaped curve as similarity decreased. That is, VWM recognition was lowest when two samples were moderately similar because the excitatory peaks of one sample were

Corresponding author: Lei Mo, Department of Psychology, South China Normal University, Zhongshan Road West, No. 55, Guangzhou, China. Phone: (+86) 20-85216892. E-mail: moleipsy@163.com

more likely to be attenuated by an inhibitory zone of another sample. In contrast, performance was good for samples that were highly similar because their excitatory peaks largely overlapped with each other and did not fall within the surrounding inhibitory zone. Furthermore, performance for stimuli with low similarity was also good, because excitatory peaks for these stimuli would fall beyond the bounds of the surrounding inhibitory zone.

Compared with low-level VWM for simple artificial objects, high-level VWM for complex real-world objects has received less attention and is not as well understood. Nevertheless, some researchers established parallels between these two domains by replicating well-known similarity effects, such as the similarity advantage effect, with real-world stimuli. Jiang, Lee, Asaad, and Remington (2016) manipulated similarity by morphing faces with a single face identity (similar condition) or multiple face identities (dissimilar condition). They found that the memory performance for the similar faces was better than that for the dissimilar faces. In contrast, in the *mixed-category benefit effect*, an increase in similarity has been shown to impair VWM. When multiple items drawn from either one or mixed categories were simultaneously presented for participants to remember, memory performance for the mixed categories was superior to that for the single category (Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014). Furthermore, the size of the mixed-category benefit was predicted by the extent to which the neural response patterns of the two categories were separated from each other within the occipito-temporal cortex (Cohen et al., 2014). These studies implied that the limits of VWM capacity in behaviour result from competition between similar representations (Franconeri, Alvarez, & Cavanagh, 2013; Wei, Wang, & Wang, 2012) and that increased similarity should result in lower VWM capacity.

Therefore, in the field of VWM for complex real-world objects, we observe a contradiction: Some studies (e.g., Jiang, Lee, et al., 2016) find that VWM improves as similarity increases, and others (e.g., Cohen et al., 2014) find the opposite results. How can the results of these studies be reconciled? One observation is that evidence supporting the similarity advantage and the mixed-category benefit used items with different degrees of similarity. Specifically, the similarity advantage was demonstrated by comparing memory for morphed items (high similarity) with nonmorphed items drawn from the same category (moderate similarity); whereas the mixed-category benefit was demonstrated by comparing memory for mixed-category nonmorphed items (low similarity) with nonmorphed items drawn from the same category (moderate similarity). It is likely that memory performance varies for items with different degrees of similarity. Memory for moderate similarity items seems to be worse than memory for either high or low similarity items. Previous research failed to reconcile these opposite similarity effects because they did not examine a breadth of similarity levels and focused on just one or two similarity levels.

To test our hypotheses, we extended the method used in Cohen et al. (2014) by adding a high similarity condition (Experiment 1). We predicted that memory performance for the moderate similarity condition would be worse than both the low similarity condition and high similarity condition. In Experiment 2, we controlled for the possibility that results were affected by stimulus familiarity differences by allowing participants to encode images for as long as they wanted. In Experiment 3, to rule out the possibility that similarity effects result from perceptual limitations rather than limitations in VWM capacity, we compared memory performance in a simultaneous presentation format with a sequential presentation format.

Our research also sheds light on the question of whether cortical resources share a common representational structure for all categories or are category-specific by investigating similarity effects. Items with high similarity are more likely to result in extraction of common properties (Lin & Luck, 2009; Sims, Jacobs, & Knill, 2012), resulting in greater within-category interference (Jiang, Remington, Asaad, Lee, & Mikkalson, 2016). One line of research reports that within-category interactions are the same across category types and, therefore, each domain-specific cortical region facilitates visual working memory in the same way (the cortical resource theory, Cohen et al., 2014). This general cortical resource theory is challenged by a memory asymmetry reported by Jiang, Remington, et al. (2016) who found that faces benefit from mixed-category presentation, but scenes do not. To investigate this divergence, we measured memory performance for faces and scenes separately.

EXPERIMENT 1

We tested whether memory performance for moderately similar objects is worse than either lowly or highly similar objects in the field of high-level vision.

Method

PARTICIPANTS

The participants were 25 undergraduate students (14 females; M_{age} = 22 years) from South China Normal University. The participants received 10 RMB in exchange for 20 min of participation. All of the participants were right-handed and had normal or corrected-to-normal vision. Each participant provided informed consent prior to his or her participation in the experiment.

DESIGN

We used a change-detection task and applied a 3 × 2 (Similarity [low, moderate, high] × Stimulus Material [faces, scenes]) within-subject design. The low similarity condition indicated nonmorphed pictures drawn from the mixed categories. The moderate similarity condition indicated nonmorphed pictures drawn from a single category. The high similarity condition indicated morphed pictures drawn from a single category (see Figure 1). The nature of the tested stimulus was defined as "face" when a face was tested in a low, moderate, or high similarity condition, and was defined as "scene" when a scene was tested in a low, moderate, or high similarity condition. The dependent variables were the Percentage of Correct Responses and Response Time (RT) in the change-detection task.
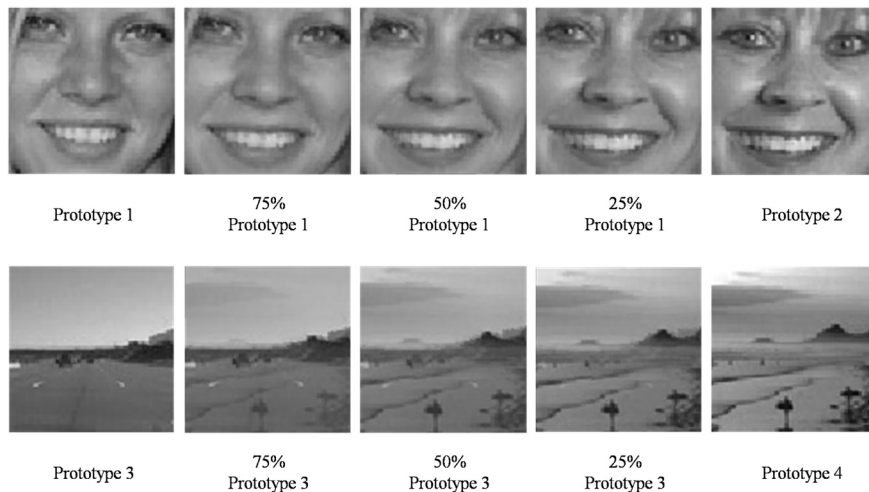
FIGURE 1.

Samples of morphed pictures used in the high similarity condition. The first line illustrates a pair of prototype faces and their morphed faces, whose similarity varied along a morph continuum. The second line illustrates a pair of prototype scenes and their morphed faces, whose similarity varied along a morph continuum.

MATERIALS AND STIMULI

We employed Cohen et al.'s (2014) stimuli, which included 40 faces and 40 scenes (see the Supporting Information section in Cohen et al., 2014). Of the 40 faces, there were 20 males and 20 females, and of the 40 scenes, there were 20 natural and 20 artificial scenes. These unaltered stimuli were used for the low and moderate similarity conditions. We used MagicMorph software (eTinysoft Inc., Shenzhen, China) to produce morphed images to be used in the high similarity conditions. This software can be used to create intermediate morphs between two different prototype images by linearly altering features (e.g., color and configuration; Freedman, Riesenhuber, Poggio, & Miller, 2001). The two prototype images in each pair were drawn from the same subcategory (e.g., scenes of lakes) and shared similar contours and key features. Three morphed images were formed from each pair of prototypes with 25%, 50%, and 75% combinations (three morphing levels, see Figure 1).

The stimulus images each subtended 6° × 6° of visual angle (6 cm × 6 cm on the computer screen; see the Procedure section below). A different image was presented in each quadrant of the visual field. Within a hemifield, the centre-to-centre distance between the items was 7.5°, and the centre-to-centre distance between two items in the different hemifields (but on the same horizontal plane) was 15.4°. A red fixation dot (0.55° × 0.55° visual angle in size) was presented in the middle of the display. The stimuli were presented on a grey background with red, green, and blue (RGB) values of 126, 126, and 126, respectively. In the low similarity condition, stimuli were drawn from two categories (two faces and two scenes); the two images from the same category were presented diagonally opposite each other (see Figure 2). In the moderate similarity condition, stimuli were drawn from the same category (i.e., four faces or four scenes) and presented in a random order across the quadrants. In the high similarity condition, stimuli were drawn from the three morphed images produced by one prototype pair and one of the two corresponding prototype images, and were presented in

a random order across the quadrants. In the test display, a red frame with a 1-pixel line width and a 6° × 6° visual angle surrounded the test item to cue its location.

The experiment was conducted on a desktop computer, and the responses were made on a keyboard with *F* and *J* as response buttons. All of the instructions and stimuli were presented on a 17 in. LCD
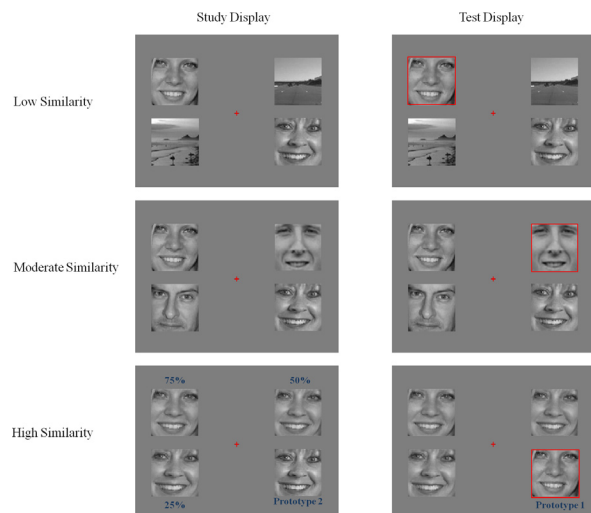
FIGURE 2.

Layout of images in study displays and test displays. In the low similarity condition, images were two original faces and two original scenes employed in the research by Cohen et al. (2014). In the moderate similarity condition, images were four original scenes or four original faces. In the high similarity condition, images were three morphed images and one of their prototype images.

monitor (1,280 × 1,024 resolution, 32-bit true colour, 75 Hz screen refresh rate). The participants sat approximately 57 cm away from the monitor, such that 1° of visual angle subtended 1 cm on the screen. The experiment was created and controlled on a computer using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, USA).

PROCEDURE

At the beginning of each trial, the red fixation dot appeared in the middle of the screen for 500 ms (see Figure 3). Then, a study display with the to-be-remembered items was presented for 800 ms. Following a second red fixation dot (1 s), the test display was presented. The participants indicated whether the item cued by the red frame in the test display was the same as the corresponding item in the study display or whether it had changed by pressing the *F* or *J* keys, respectively, as quickly and as accurately as possible. Participants were first required to complete practice trials until they reached a correct response rate of 75%.

In the main experiment, the participants completed six blocks of 20 trials (120 total experimental trials), which yielded a total of 40 low similarity trials, 40 moderate similarity trials, and 40 high similarity trials. Within each block, all trial types appeared in a random order. For each participant, the test item was the same as the item in the corresponding location of the study display for half of the 120 trials, but was changed for the other half. Whenever an item changed, it did so to another item from the same category (e.g., a face might change into another face but not a scene), and the change occurred only within the cued location. In both the low similarity and moderate similarity conditions, half of the changed trials involved a switch between subcategories (e.g., a change from a male to female image in a face condition or from natural to artificial in a scene condition). In the high similarity condition, the probed image was always the prototype, which changed into the other prototype image from the same prototype pair. This method of manipulating study-test item similarity is consistent with Experiment 2 by Jiang, Remington, et. al. (2016), which increased the similarity among scenes by employing scenes drawn from the same subcategory.

Results

We performed a 3 × 2 repeated-measures analysis of variance (ANOVA) with Percentage of Correct Responses as the dependent variable and Similarity (low, moderate, high) and Stimulus Materials (faces, scenes) as factors. Consistent with our hypothesis of similarity effects, we found a significant main effect of similarity, $F(2, 48) = 49.26$, $p < .001$, $\eta^2_p = .67$ (see Figure 4). Post hoc tests indicated that accuracy was significantly higher in the low similarity condition (74.8%) than the moderate similarity condition (65.7%), $F(1, 24) = 17.63$, $p < .001$, $\eta^2_p = .42$, and higher in high similarity condition (85.1%) than the moderate similarity condition (65.7%), $F(1, 24) = 89.91$, $p < .001$, $\eta^2_p = .79$. The similarity effects differed for faces and scenes, yielding a significant two-way interaction between similarity and stimulus materials, $F(2, 48) = 3.81$, $p = .029$, $\eta^2_p = .14$. The other effects were not significant.

To understand the interaction among similarity and stimulus materials, we separately analyzed data for faces and scenes. The pattern for faces alone was consistent with the overall similarity effect. We found a significant main effect of similarity, $F(2, 48) = 18.34$, $p < .001$, $\eta^2_p = .43$. Memory for faces in the low similarity condition (78.4%) was significantly better than that in the moderate similarity condition (68.4%), $F(1, 24) = 11.79$, $p = .002$, $\eta^2_p = .33$. This result indicates that memory for faces has a mixed-category benefit. Memory for faces in the high similarity condition (83.8%) was significantly better than in the moderate similarity condition (68.4%), $F(1, 24) = 50.52$, $p < .001$, $\eta^2_p = .68$. This result indicates that memory for faces has a similarity benefit. In addition, memory for faces in the low similarity condition (78.4%) was significantly worse than in the high similarity condition (83.8%), $F(1, 24) = 4.32$, $p = .049$, $\eta^2_p = .15$.

Memory for scenes was also consistent with similarity effects. We found a significant main effect of similarity, $F(2, 48) = 35.08$, $p < .001$, $\eta^2_p = .59$. Memory for scenes in the low similarity condition (71.2%) was significantly better than in the moderate similarity condition (63.1%), $F(1, 24) = 6.64$, $p = .017$, $\eta^2_p = .22$. This result indicates that memory for scenes has a mixed-category benefit. Memory for scenes in high similarity condition (86.4%) was significantly better than in the moderate similarity condition (63.1%), $F(1, 24) = 62.49$, $p < .001$, $\eta^2_p = .72$. This result indicates that memory for scenes has a similarity advantage. In addition, memory for scenes in the low similarity condition (71.20%) was significantly worse than in the high similarity condition (86.40%), $F(1, 24) = 43.68$, $p < .001$, $\eta^2_p = .65$.

To further understand the interaction among similarity and stimulus materials, we also separately compared memory for faces and scenes within the low, moderate, and high similarity conditions. In the low similarity condition, memory for faces (78.4%) was not significantly different from that for scenes (71.2%), $F(1, 24) = 4.26$, $p =$
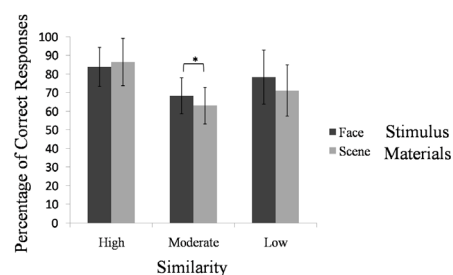
FIGURE 3.

Procedure of a change-detection task. A study display and a test display were presented successively. In each display, four items were presented simultaneously. Participants were required to indicate whether the cued item in the test display was the same as the corresponding item in the study display.

FIGURE 4.

Average memory accuracy for faces and scenes in different similarity conditions in Experiment 1. Error bars reflect *SDs*.

.050, η^2_p = .15. In the moderate similarity condition, memory for faces (68.4%) was significantly better than that for scenes (63.1%), $F(1, 24)$ = 6.57, p = .017, η^2_p = .22. In the high similarity condition, memory for faces (83.8%) was not significantly different from that for scenes (86.4%), $F(1, 24)$ = 1.03, p = .321, η^2_p = .04. These results indicated that the mixed-category benefit for scenes was more profound than for faces. Similarly, the similarity advantage for scenes was also more profound than for faces.

We also performed a repeated-measures ANOVA on RT using Similarity and Stimulus Materials as factors. We found a significant main effect of similarity, $F(2, 48)$ = 8.87, p = .001, η^2_p = .27. A further multiple comparison analysis revealed that RT in the low similarity condition (1,108 ms) was faster than that in the moderate similarity condition (1,168 ms), $F(1, 24)$ = 5.55, p = .027, η^2_p = .19. RT in the high similarity condition (1,060 ms) was faster than that in the moderate similarity condition (1,168 ms), $F(1, 24)$ = 22.00, p < .001, η^2_p = .48. RT in the low similarity condition (1,108 ms) was similar to that in the high similarity condition (1,060 ms), $F(1, 24)$ = 2.88, p = .103, η^2_p = .11. These parallel results for accuracy and RT indicated that participants did not trade off speed and accuracy. We also found a significant main effect of stimulus materials, $F(1, 24)$ = 6.65, p = .016, η^2_p = .22.

Discussion

One important finding emerging from Experiment 1 was that both memory for faces and scenes showed similarity effects. Although these two categories differed in the strength of similarity effects, the similarity effects were evident in both of them.

One potential explanation of the similarity effects lies in differences in familiarity among items presented simultaneously. It is possible that high memory performance for the prototype images in the high similarity condition could have resulted from the fact that these stimuli were also presented to participants in the low and moderate similarity conditions, whereas the morphed images used in the high similarity condition were not presented in the other two conditions. Participants may have paid more attention to the familiar prototype images than the morphed images (Christie & Klein, 1995), which may have facilitated perceptual processing and resulted in improved performance for prototype images in high similarity condition (Hawkins et al., 1990). In contrast, in the low and moderate similarity conditions, participants would have paid equal attention to four images with equal familiarity, resulting in inferior memory performance. To test the possibility that similarity effects were due to differences in stimulus familiarity, in Experiment 2, we allowed participants to encode items for as long as they wanted. If familiarity differences primarily caused the similarity advantage, then it should disappear in self-paced encoding. However, if memory capacity limitations are the primary cause of the similarity advantage, then the effect should still occur in self-paced encoding.

EXPERIMENT 2

To exclude the possibility that uneven attention allocation caused by familiarity differences contributed to the similarity effects, we allowed participants to encode items for as long as they wanted. If the similarity effects benefit from familiarity differences, these effects should be attenuated by self-paced encoding. If similarity effects primarily result from memory capacity limitations, these effects should be maintained in self-paced encoding.

Method

PARTICIPANTS.

The participants were 36 students (33 females; M_{age} = 20 years) from South China Normal University. The participants received 10 RMB in exchange for 20 min of participation. All of the participants were right-handed and had normal or corrected-to-normal vision. Each participant provided informed consent prior to his or her participation in the experiment.

DESIGN AND PROCEDURE.

We applied a 3 × 2 (Similarity [high, moderate, low] × Stimulus Type [face, scene]) within-subject design. The dependent variables were Encoding Time, Accuracy Rate, and RT in the change-detection task. Equipment and stimuli were the same as those in Experiment 1. The procedure was the same as in Experiment 1, except for the change to the encoding procedure. Instead of viewing the study display for a fixed amount of time, participants were allowed to view the study display as long as they wanted. They indicated that they had finished encoding by pressing the spacebar.

Results

The amount of encoding time participants took (M = 3,696 ms, SE = 2,116 ms) was longer than the limited encoding time (800 ms) in Experiment 1. Encoding time differed significantly across similarity, $F(2, 70)$ = 14.75, p < .001, η^2_p = .30. A further multiple comparison analysis revealed that encoding time for low similarity images (3,117 ms) was shorter than moderate similarity images (3,853 ms), $F(1, 35)$ = 19.40, p < .001, η^2_p = .36, and encoding time for low similarity images (3,117 ms) was also shorter than for high similarity images (4,117 ms), $F(1, 35)$ = 15.77, p < .001, η^2_p = .31. Encoding time for high and moderate similarity images were not significantly different, $F(1, 35)$ = 3.88, p = .057, η^2_p = .10.

We performed a repeated-measures ANOVA on percentage of correct responses using Similarity and Stimulus Types as factors. Consistent with the similarity effects found in Experiment 1, we found a significant main effect of similarity, $F(2, 70)$ = 119.10, p < .001, η^2_p = .77, with higher accuracy in low similarity images (92.8%) than moderate similarity images (77.5%), $F(1, 35)$ = 174.57, p < .001, η^2_p = .83, and higher accuracy in high similarity images (94.4%) than moderate similarity images (77.5%), $F(1, 35)$ = 146.09, p < .001, η^2_p = .81 (see Figure 5). Since participants had enough time to encode every image of a display, any perceptual limitations caused by familiarity differences were eliminated. Therefore, this result helps exclude the possibility that it is a perceptual limitation rather than VWM limitation that results in the similarity effects. The other effects were not significant.

We performed a repeated-measures ANOVA on RT using Similarity and Stimulus Type as factors. We found a significant main effect of
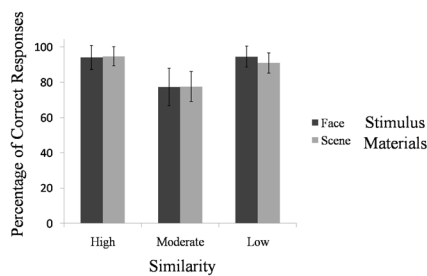
FIGURE 5.

Average memory accuracy for faces and scenes in different similarity conditions in Experiment 2. Error bars reflect *SD*s.

similarity, $F(2, 70) = 17.58$, $p < .001$, $\eta^2_p = .33$. We also found a significant interaction between similarity and stimulus type, $F(2, 70) = 8.30$, $p = .001$, $\eta^2_p = .19$. Post hoc analyses revealed that RT for low similarity images (1,268 ms) was shorter than moderate similarity images (1,501 ms), $F(1, 35) = 38.48$, $p < .001$, $\eta^2_p = .52$. RT for high similarity images (1,357 ms) was shorter than moderate similarity images (1,501 ms), $F(1, 35) = 11.24$, $p = .002$, $\eta^2_p = .24$. In addition, RT for low similarity images (1,268 ms) was shorter than high similarity images (1,357 ms), $F(1, 35) = 5.44$, $p = .026$, $\eta^2_p = .13$. These results indicate that, as in Experiment 1, participants did not show a speed-accuracy trade-off.

Discussion

Experiment 2 asked participants to encode images for as long as they wanted, ensuring that uneven attention allocation caused by a familiarity difference was abolished. Nonetheless, we continued to find similarity effects for both faces and scenes. Unlike Experiment 1, we did not find memory asymmetry here between faces and scenes within high, moderate, and low similarity conditions. Given that accuracy for both high similarity items and low similarity items was over 90%, this lack of memory asymmetry might be attributed to ceiling effects.

Since all stimuli were presented simultaneously for encoding, another possibility is that similarity effects could result from perceptual functions, rather than memory. In Experiment 3, we aimed to reduce the role of perception by presenting images sequentially.

EXPERIMENT 3

In order to test whether similarity effects are due to perceptual or working memory processing, we compared similarity effects when items were presented either simultaneously or sequentially. When items are presented simultaneously, as in Experiment 1, they must share perceptual encoding capacity and compete with one another. However, when items are presented sequentially, their perceptual encoding will not compete because only a single item is presented at any given time. Therefore, if the similarity effect is caused by interference exclusively during perceptual encoding, we would expect to see a similarity effect only in the simultaneous condition. If the similarity effects arise due to VWM capacity limitations, similarity effects should be observed for both simultaneous and sequential presentation formats.

Method

PARTICIPANTS

Thirty participants (4 females; M_{age} = 19 years) were recruited from South China Normal University. The participants received 20 RMB for completing the 30 min session. All of the participants were right-handed and had normal or corrected-to-normal vision. Each participant provided informed consent prior to his or her participation in the experiment.

DESIGN

We utilized a 3 × 2 × 2 (Similarity [high, moderate, low] × Stimulus Type [face, scene] × Presentation [simultaneous, sequential]) within-subject design. The dependent variables were Accuracy Rate and RT in a change-detection task.

Equipment and stimuli were the same as those in Experiment 1.

PROCEDURE

In the sequential condition, each item was initially presented for 200 ms. The total encoding time was 800 ms, enabling the sequential condition to be directly comparable with the simultaneous condition. A 1 s delay separated the presentation of the last encoding item and the test display, which only contained a single item as in the simultaneous condition. The simultaneous condition unavoidably results in differences in delay between items and probes depending on the sequential order: For example, if the first item presented was probed, the item-probe delay would be 1.6 s, whereas if the last item was probed, the item-probe delay was only 1.s. In order to control for any effect of delay length, we varied the item-probe delays in the simultaneous condition across the same range. Furthermore, we probed stimuli at each temporal position (first, second, third, or fourth) an equal number of times so that participants could not use temporal information to predict which item would be probed.

The procedure in the simultaneous condition was similar to Experiment 1, with the following exceptions. First, between the presentation of the four stimuli and the probe, the retention interval varied in duration across trials to match the retention intervals for the sequential displays. Specifically, the retention intervals were 1.6 s, 1.4 s, 1.2 s, and 1 s (see Figure 6). This manipulation ensured that the average retention duration for the probed item was matched between the two conditions. Second, in the test display, only one item was presented without a red frame as a cue. Participants indicated whether that particular item changed by pressing the *F* or *J* keys.

Results

We performed a repeated-measures ANOVA on percentage of correct responses with Similarity, Stimulus Materials, and Presentation as factors. Consistent with the similarity effects reported in the previous studies, we found a significant main effect of similarity, $F(2, 58) = 74.10$, $p < .001$, $\eta^2_p = .72$, with higher accuracy for high similarity items (85.1%) than moderate similarity items (70.0%), $F(1, 29) = 112.71$, $p < .001$, $\eta^2_p = .80$, and with higher accuracy for low similarity items

Sequential   **Simultaneous**
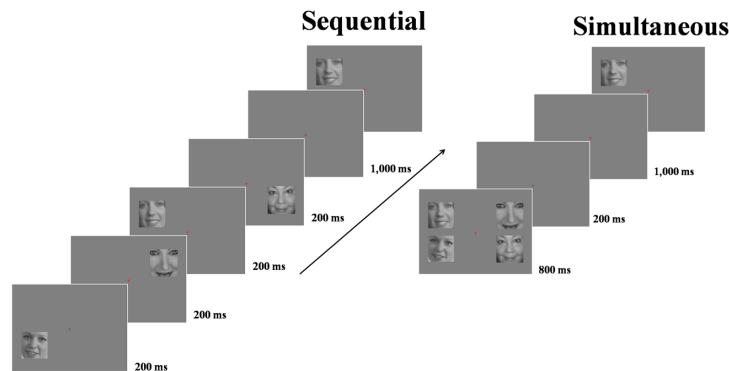
FIGURE 6.

Sequential presentation task. On each trial, four images were presented at each quadrant of visual field at a pace of 200 ms per image. After an interval of 1 s, a test image was presented for participants to make a "same /different" decision. In the simultaneous condition, item-probe delays varied across the same range as those in the sequential condition.

(78.67%) than moderate similarity items (70%), $F(1, 29) = 90.26$, $p < .001$, $\eta_p^2 = .76$ (see Figure 7). These similarity effects were found in both presentation formats, resulting in a lack of significant two-way interaction between similarity and presentation, $F(2, 58) = 1.86$, $p = .165$, $\eta_p^2 = .06$. This result indicated that similarity effects arise at least in part because of VWM capacity limitations rather than perceptual encoding limitations.

We also found a significant main effect of presentation, $F(1, 29) = 7.94$, $p = .009$, $\eta_p^2 = .22$. Further multiple comparison analysis revealed that memory for items presented simultaneously (80.3%) was significantly better than sequentially (75.6%), $F(1, 29) = 7.94$, $p = .009$, $\eta_p^2 = .22$.

Furthermore, the similarity effects differed between faces and scenes, yielding a significant two-way interaction between similarity and stimulus materials, $F(2, 58) = 10.52$, $p < .001$, $\eta_p^2 = .27$. To understand the interaction between similarity and stimulus materials, we separately analyzed data for faces and scenes.

Memory for faces was consistent with similarity effects. We found a significant main effect of similarity, $F(2, 58) = 21.41$, $p < .001$, $\eta_p^2 = .43$. Memory for high similarity items (82.7%) was significantly better than moderate similarity items (71.4%), $F(1, 29) = 37.21$, $p < .001$, $\eta_p^2 = .56$. Memory for low similarity items (81.2%) was significantly better than moderate similarity items (71.42%), $F(1, 29) = 29.23$, $p < .001$, $\eta_p^2 = .50$. These results indicated similarity effects in memory for faces

and suggested that a centre-surround inhibition format might underlie face memory.

Memory for scenes was also consistent with similarity effects. We found a significant main effect of similarity, $F(2, 58) = 84.50$, $p < .001$, $\eta_p^2 = .74$. Memory for items with high similarity (87.5%) was significantly better than for moderate similarity (68.6%), $F(1, 29) = 152.16$, $p < .001$, $\eta_p^2 = .84$. Memory for items with low similarity (76.2%) was significantly better than for moderate similarity (68.6%), $F(1, 29) = 27.70$, $p < .001$, $\eta_p^2 = .49$. These results indicated similarity effects in memory for scenes and suggested that a centre-surround inhibition format might underlie scene memory.

To further understand the interaction between similarity and stimulus materials, we separately analyzed data in low, moderate, and high similarity conditions. In the low similarity condition, memory for faces (81.3%) was significantly better than that for scenes (76.2%), $F(1, 29) = 5.51$, $p = .026$, $\eta_p^2 = .16$. In the moderate similarity condition, memory for faces (71.4%) was not significantly different from that for scenes (68.6%), $F(1, 29) = 3.25$, $p = .082$, $\eta_p^2 = .10$. In the high similarity condition, memory for faces (83.7%) was significantly worse than that for scenes (87.5%), $F(1, 29) = 14.93$, $p = .001$, $\eta_p^2 = .34$. Combining the results in the low and moderate similarity conditions, we could draw the conclusion that the mixed-category benefit for faces was more profound than for scenes. Combining the results in the high and moderate similarity conditions, we could draw the conclusion that the similarity advantage for scenes was more profound than for faces.

We also performed a repeated-measures ANOVA on RT using Similarity, Stimulus Materials, and Presentation as factors. We found a significant main effect of similarity, $F(2, 58) = 13.46$, $p < .001$, $\eta_p^2 = .32$. RT for items in the high similarity condition (872 ms) was faster than in the moderate similarity condition (956 ms), $F(1, 29) = 30.41$, $p < .001$, $\eta_p^2 = .51$. RT for items in the low similarity condition (902 ms) was faster than in the moderate similarity condition (956 ms), $F(1, 29) = 7.03$, $p = .013$, $\eta_p^2 = .20$. RT for items in the high similarity condition (872 ms) was faster than in the low similarity condition (902 ms), $F(1, 29) = 5.56$, $p = .025$, $\eta_p^2 = .16$. These results indicate that participants did not show a speed-accuracy trade-off.
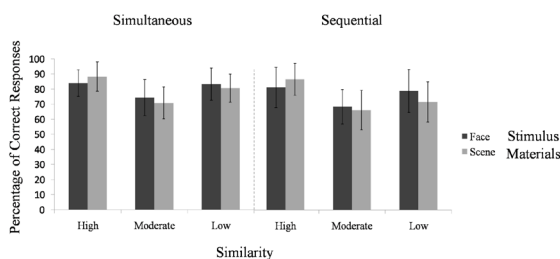
FIGURE 7.

Average memory accuracy for faces and scenes in different similarity conditions in Experiment 3. Error bars reflect *SD*s.

Discussion

Experiment 3 rules out one possible explanation that similarity effects are due to perceptual encoding limitations. Instead, similarity effects occurred, at least in part, during VWM maintenance. We found a mixed-category benefit and a similarity advantage for both faces and scenes, but also noticed a memory asymmetry: The mixed-category benefit was driven more by faces than scenes, whereas the similarity advantage was driven more by scenes than faces. As discussed in the General Discussion section, these data have implications for our theoretical understanding of memory asymmetry.

It is also worth noting that there was a main effect of presentation format with working memory performance in the simultaneous conditions being better than in the sequential conditions. Given that presenting items sequentially eliminates attentional limitations and perceptual competition among stimulus representations (Ihssen, Linden, & Shapiro, 2010), our results further support the notion that inhibition between items occurs during working memory maintenance.

GENERAL DISCUSSION

We tested how similarity modulates memory performance for high-level objects. We found memory for moderate similarity items was worse than for either low or high similarity items. These similarity effects suggested that high-level VWM might be characterized by centre-surround inhibition. We also showed that similarity effects were not equally strong for faces and scenes, supporting category-specific cortical resource theories.

We found that similarity effects were robust. This phenomenon held regardless of the categories of stimuli (faces or scenes), encoding duration (limited or self-paced), or presentation mode (simultaneous or sequential). Based on the evidence for centre-surround inhibition in low-level vision (Störmer & Alvarez, 2014), the opposing similarity effects in high-level VWM representation might be produced by an analogous centre-surround selection mechanism. Specifically, high similarity items received the greatest processing benefit because they might all fall within the excitatory centre, whereas processing of moderate similarity items suffered because they might fall in the range of the inhibitory surround. High dissimilarity items recovered because they may be at a more distant location outside the range of the inhibitory surround.

However, an alternative possibility of strong memory performance for high similarity items is a reduced memory load. On the one hand, the familiarity difference between prototypes and morphed images made it possible for participants to select the more familiar image to encode. On the other hand, given morphed images were reported as less natural than prototypes, it is likely that participants selected more natural images to encode. These potential strategies could effectively reduce the memory load to only one item, resulting in the best memory performance for high similarity items. These possibilities cannot be examined by the present behavioral data, but might be examined by contralateral delay activity (CDA), an EEG component the amplitude

of which reflects the number of items stored in VWM (Luck & Vogel, 2013; Vogel & Machizawa, 2004). Future work could compare CDA amplitudes to investigate this issue.

The strong memory performance for high similarity stimuli is unlikely to be due to differences in sample-test similarity among conditions. It is true that the average sample-test similarity in the high similarity condition was slightly higher than in the low and moderate similarity conditions, but, if anything, this should have hurt rather than helped accuracy. According to research by Awh, Barton, and Vogel (2007), increase in sample-test similarity results in worse performance in a change-detection task. Even so, we believe that future research should strictly control the magnitude of change across different similarity conditions.

Although similarity effects were overall advantageous to VWM performance, we found that these effects were not equally strong for different categories of stimuli, at least for faces and scenes. On the one hand, we found the mixed-category benefit for faces was stronger than that for scenes. This result was consistent with the research by Jiang, Remington, et al. (2016) and supported the category-specific cortical resources theory. On the other hand, we also found the similarity advantage for scenes was stronger than for faces. These results extend the content of *memory asymmetry* which was first proposed by Jiang, Remington, et al. (2016).

According to these results, we found that the centre-surround inhibition format for high-level VWM might be different from that for attention or low-level VWM, because the centre-surround inhibition format, well-established in the field of attention and low-level VWM, cannot explain the memory asymmetry phenomenon observed here. We speculate that different categories may have different centre-surround inhibition organizations with different sizes of the excitatory peak centre and the surrounding inhibitory zone. Specifically, faces may be represented with relatively small excitatory peak centres and larger surrounding inhibitory zones, whereas scenes may be represented with larger excitatory peak centres and smaller sizes of inhibitory zones (see Figure 8). These differences in excitatory and inhibitory region size could account for both the mixed-category and similarity effects. When faces are encoded with other faces, their excitatory centres would largely fall into inhibitory zones of other faces (see Figure 8, moderate similarity condition for faces), but when faces are encoded with scenes, their excitatory centres would be largely separated from inhibitory zones of other items, resulting in a large increase in memory performance for faces (see Figure 8, low similarity condition). When scenes are encoded with other scenes, their excitatory centres would be unlikely to fall into the inhibitory zones of other items (see Figure 8, moderate similarity condition for scenes), resulting in little increase in memory performance for scenes when scenes are encoded with faces. These patterns together would result in a mixed-category benefit for faces greater than that for scenes consistent with our results. Furthermore, when scenes sharing more common properties are encoded together, their excitatory zones would be more likely to overlap, resulting in less suppressive interaction (Beck & Kastner, 2007), stronger activation, and better memory performance (see Figure 8, high similarity condi-
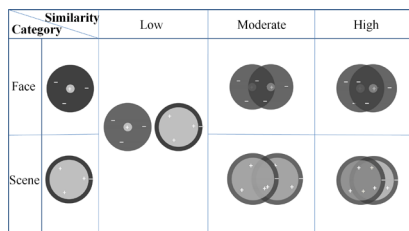
FIGURE 8.

Idealized schematic depiction of the memory asymmetry phenomenon between faces and scenes. Faces may have centre-surround inhibition organization with relatively small excitatory peak centres and larger surrounding inhibitory zones. Scenes may have centre-surround inhibition organization with larger excitatory peak centres and smaller size of inhibitory zones. These differences may result in the fact that faces benefit more from mixed-category encoding than scenes, whereas scenes benefit more from similarity advantage than faces.

tion for scenes). In contrast, when faces sharing more properties are encoded together, their excitatory zones would be less likely to overlap with each other, resulting in little increase in memory performance for faces (see Figure 8, high similarity condition for faces). These effects together would account for our findings that the similarity advantage for scenes is stronger than that for faces. This explanation is merely speculation and future research should be carried out to examine this issue.

CONCLUSIONS

Similarity has been observed to have opposite effects on VWM in different studies. This contradiction was reconciled through manipulations of similarity and comparing memory performance across a broader range of similarity levels. The mechanism that underlies this phenomenon might be high-level centre-surround inhibition. Besides typical characteristics of centre-surround inhibition, such as an excitatory peak centre and a surrounded inhibitory zone, we propose that high-level centre-surround inhibition organization has category-specific sizes of the excitatory peak centre and surrounding inhibitory zone. These category-specific differences result in unequally strong effects of similarity.

AUTHOR NOTE

Li Yang is now at Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, China, yang_li_16@163.com; Lei Mo is now at Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, China, moleipsy@163.com

Authors have no conflicts of interest to disclose.

REFERENCES

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*, 622–628. doi: 10.1111/j.1467-9280.2007.01949.x

Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision, 7*:19. doi: 10.1167/7.2.19

Christie, J., & Klein, R. (1995). Familiarity and attention: Does what we know affect what we notice? *Memory & Cognition, 23*, 547–550. doi: 10.3758/BF03197256

Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 8955–8960. doi: 10.1073/pnas.1317860111

Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: Psychophysical evidence for a suppressive annulus around an attended item. *Vision Research, 43*, 205–219. doi: 10.1016/S0042-6989(02)00491-1

E-Prime 2.0 [Computer software]. Pittsburgh, USA: Psychology Software Tools.

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences, 17*, 134–141. doi: 10.1016/j.tics.2013.01.010

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001, January). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*(5502), 312–316. doi: 10.1126/science.291.5502.312

Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 802–811. doi: 10.1037/0096-1523.16.4.802

Hopf, J.-M., Boehler, C., Luck, S., Tsotsos, J., Heinze, H.-J., & Schoenfeld, M. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 1053–1058. doi: 10.1073/pnas.0507746103

Ihssen, N., Linden, D. E., & Shapiro, K. L. (2010). Improving visual short-term memory by sequencing the stimulus array.

Psychonomic Bulletin & Review, 17, 680–686. doi: 10.3758/PBR.17.5.680

Jiang, Y. V., Lee, H. J., Asaad, A., & Remington, R. (2016). Similarity effects in visual working memory. *Psychonomic Bulletin & Review, 23*, 476–482. doi: 10.3758/s13423-015-0905-5

Jiang, Y. V., Remington, R. W., Asaad, A., Lee, H. J., & Mikkalson, T. C. (2016). Remembering faces and scenes: The mixed-category advantage in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1399–1411. doi: 10.1037/xhp0000228

Kiyonaga, A., & Egner, T. (2016). Centre-surround inhibition in working memory. *Current Biology, 26*, 64–68. doi: 10.1016/j.cub.2015.11.013

Lin, P.-H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition, 17*, 356–372. doi: 10.1080/13506280701766313

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*, 391–400. doi: 10.1016/j.tics.2013.06.006

MagicMorph software [Computer software]. Shenzhen, China: eTinysoft Inc.

Mate, J., & Baqués, J. (2009). Visual similarity at encoding and retrieval in an item recognition task. *Quarterly Journal of Experimental Psychology, 62*, 1277–1284. doi: 10.1080/17470210802680769

Müller, N. G., Mollenhauer, M., Rösler, A., & Kleinschmidt, A. (2005). The attentional field has a Mexican hat distribution. *Vision Research, 45*, 1129–1137. doi: 10.1016/j.visres.2004.11.003

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119*, 807–830. doi: 10.1037/a0029856

Störmer, V. S., & Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology, 24*, 1985–1988. doi: 10.1016/j.cub.2014.07.030

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748–751. doi: 10.1038/nature04171

Wei, Z., Wang, X.-J., & Wang, D.-H. (2012). From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *The Journal of Neuroscience, 32*, 11228–11240. doi: 10.1523/JNEUROSCI.0735-12.2012