CrossMark

BRIEF REPORT

Similarity effects in visual working memory

Yuhong V. Jiang[1] · Hyejin J. Lee[1] · Anthony Asaad[1] · Roger Remington[1]

Abstract Perceptual similarity is an important property of multiple stimuli. Its computation supports a wide range of cognitive functions, including reasoning, categorization, and memory recognition. It is important, therefore, to determine why previous research has found conflicting effects of inter-item similarity on visual working memory. Studies reporting a similarity advantage have used simple stimuli whose similarity varied along a featural continuum. Studies reporting a similarity disadvantage have used complex stimuli from either a single or multiple categories. To elucidate stimulus conditions for similarity effects in visual working memory, we tested memory for complex stimuli (faces) whose similarity varied along a morph continuum. Participants encoded 3 morphs generated from a single face identity in the similar condition, or 3 morphs generated from different face identities in the dissimilar condition. After a brief delay, a test face appeared at one of the encoding locations for participants to make a same/different judgment. Two experiments showed that similarity enhanced memory accuracy without changing the response criterion. These findings support previous computational models that incorporate featural variance as a component of working memory load. They delineate limitations of models that emphasize cortical resources or response decisions.

✉ Yuhong V. Jiang
jiang166@umn.edu

[1] Department of Psychology, University of Minnesota, 75 East River Road, S251 Elliott Hall, Minneapolis, MN 55455, USA

Daily activities such as crossing a busy street or remembering the locations of conversation partners require us to maintain and manipulate visual information briefly in mind. The system that supports this function – visual working memory – correlates with general intelligence and has been under intense investigation (Luck & Hollingworth, 2008). Scientific debates have centered primarily around the nature of the capacity limitation (Suchow, Fougnie, Brady, & Alvarez, 2014). Recently, research has increasingly focused on inter-item interactions in visual working memory. Inter-item properties include ensemble or summary statistics of multiple items (e.g., darker objects are clustered in one area; Brady & Alvarez, 2014), the spatial configuration of memory items (Jiang, Olson, & Chun, 2000), and perceptual grouping of objects (Woodman, Vecera, & Luck, 2003). The present study examines one type of inter-item relationship: perceptual similarity among items held in working memory.

Does perceptual similarity among multiple items facilitate or hinder our ability to retain and retrieve information in visual working memory? This question has received contradictory answers. One pattern of data suggests that similarity is detrimental. For example, memory for an array of faces and scenes is better than memory for just faces or scenes (Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014). This finding is consistent with those from verbal working memory, which show better memory for phonologically dissimilar than similar words (Baddeley, 1966). They support theories such as the multiple-resource theory (Olson & Jiang, 2002; Wheeler & Treisman, 2002) or its neural instantiation – the cortical resource theory (Cohen et al., 2014). Dissimilar stimuli yield better memory because peaks of neural activation for these stimuli are widely separated, minimizing interference.

Conversely, other studies have found that similarity among items improves visual working memory. Lin and Luck (2009) showed that participants were better at detecting subtle color changes after encoding three highly similar colors (e.g., three

shades of red) than three dissimilar colors (e.g., red, green, and yellow). This pattern also emerged in visual working memory for line orientation and length (Sims, Jacobs, & Knill, 2012). Though counterintuitive, the finding fits with several computational models. For example, in information theoretic accounts, variance in the encoded features is a source of noise that interferes with encoding. High variance among memory items reduces the ability of a limited-capacity system in maintaining precise representations (Sims et al., 2012).

How might these empirical outcomes be reconciled? One observation is that they result from the use of different stimulus materials. Studies finding an advantage for dissimilar items primarily used stimuli drawn from separate perceptual and conceptual categories, whereas studies showing an advantage for similar items have used stimuli within a more restricted range of similarity. It is possible that when stimuli come from a narrower region of similarity space (e.g., different shades of red), participants are better able to extract relative features along the similarity dimension. In contrast, highly distinct stimuli from a single category (e.g., different scenes) do not support the extraction of common perceptual values and therefore show no memory benefit relative to mixed-category stimuli. However, the studies to date have largely conflated categorical distinctiveness with stimulus complexity. Those that varied similarity in a confined featural space have primarily used simple stimuli, such as colors and orientations, whereas those that contrasted single and mixed categories have used complex stimuli, such as faces and scenes. It is therefore difficult to determine the conditions that yield the two patterns of results. The current study resolves this ambiguity by testing complex stimuli that vary along a morph continuum.

In two experiments we evaluated the effects of visual similarity using faces. Participants encoded three face morphs derived from either the same individual (the similar condition) or different individuals (the dissimilar condition). Previous research using fMRI showed that following the presentation of one face identity, the fusiform face area responded less strongly to face morph of the same perceived identity relative to that of a different perceived identity (Rotshtein, Henson, Treves, Driver, & Dolan, 2005). Assuming that repetition suppression occurs when the same neural population is engaged repeatedly, the fMRI finding suggests that different face identities rely on more diverse neural populations than the same identity. If complex stimuli benefit from increased cortical resources, memory for dissimilar faces should be superior to memory for similar faces. In contrast, if reduced variance lowers working memory load for similar faces (Sims et al., 2012), then memory for similar faces should be superior to memory for dissimilar faces.

Our study also sheds light on a third pattern of findings in the literature: similarity influences response criterion but not memory sensitivity. In a series of studies Sekuler, Kahana, and colleagues have consistently reported effects of inter-item similarity on the criterion that participants used to make the

"same/different" judgment. Participants viewed a sequence of visual stimuli at fixation, followed by a test probe (Kahana, Zhou, Geller, & Sekuler, 2007; Nosofsky & Kantner, 2006; Viswanathan, Perl, Visscher, Kahana, & Sekuler, 2010). The task was to determine whether the test probe was the same as one of the encoded stimuli or different from all of them. This work revealed two mechanisms. First, the same/different decision is derived on the basis of global similarity between the test probe and all memory items, a process known as "ensemble coding" (Alvarez, 2011). Second, similarity among the memory items produces a homogeneity signal. This signal adjusts the old/new decision criterion such that participants are less likely to respond "same" if the homogeneity signal is high. Because high similarity reduces false alarms at the same time as it depresses hit rates, it affects response criterion rather than memory sensitivity. Kahana and Sekuler observed this pattern using both orientations and faces, yet they always presented memory stimuli serially in the same location. Here we examined whether Kahana and Sekuler's pattern of results extended to tasks that presented stimuli in different locations.

Method

Participants Twenty college students participated in Experiment 1, and 20 others participated in Experiment 2. They were 18 to 35 years old, naïve to the purpose of the study, and had normal or corrected-to-normal visual acuity.

Equipment Participants were tested individually in a room with normal interior lighting. Stimuli were generated using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997) implemented in MATLAB (www.mathworks.com) and displayed on a 17-in. CRT monitor (1,024 × 768 pixels; 75 Hz). Viewing distance was approximately 40 cm.

Materials We created prototype faces of eight Caucasian male celebrities using FaceGen Modeller (www.facegen.com). The program removed external features such as hair. All faces were front view and had neutral or slightly smiling expressions. We morphed each prototype with FaceGen's average male face. Each prototype had eight levels of morph, containing 30 % to 100 % of the original prototype, in steps of 10 % (Jiang, Shim, & Makovski, 2008). The total stimulus set comprised 64 faces, eight morphs from each of eight prototypes (see Fig. 1). Each face subtended 9.8° × 9.8° and was in gray scale. Faces were displayed against a black background.

Encoding displays were of two types: similar faces and dissimilar faces. On each trial we randomly chose three different morph levels (e.g., 30 %, 50 %, and 70 %). These levels were assigned to a randomly selected prototype (e.g., Face A) in the similar condition (e.g., A30, A50, and A70), or to three randomly selected prototypes
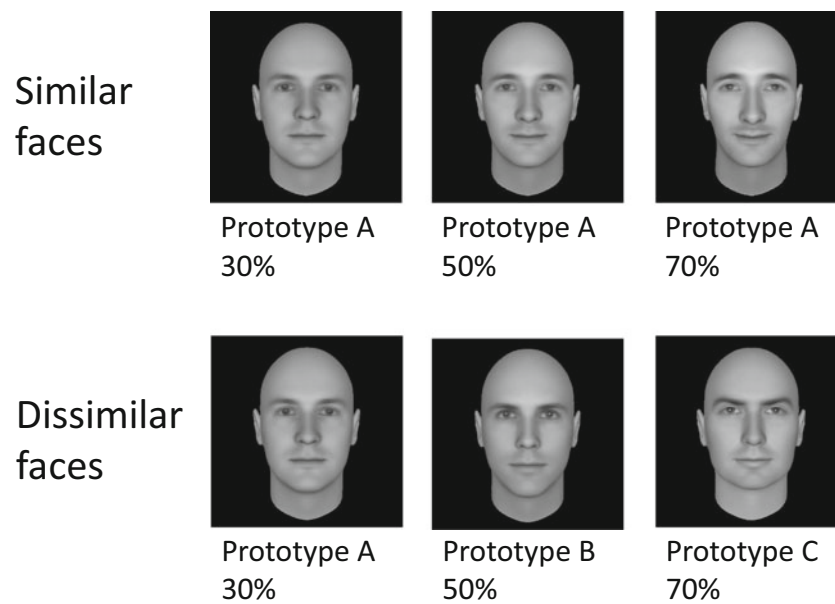
Fig. 1 Sample stimuli used in the similar and dissimilar conditions

in the dissimilar condition (e.g., A30, B50, and C70). No two faces on a given memory display were the same. The three faces were presented at equidistant locations on an imaginary circle with a radius of 7.4° (see Fig. 2).

The probe display contained a single face that appeared at the same location as one of the memory faces. The probe was either the same as the face in that location before, or a different

morph from the same prototype. When a different face was presented, it always differed from the target memory by 40 %. For example, suppose the original faces were A30, B50, and C70, and the probe face appeared in the location of B50. On same trials the probe would be B50. On different trials the probe would be B90. If the probe had appeared in the location of C70, it could be either C70 (same) or C30 (different). We
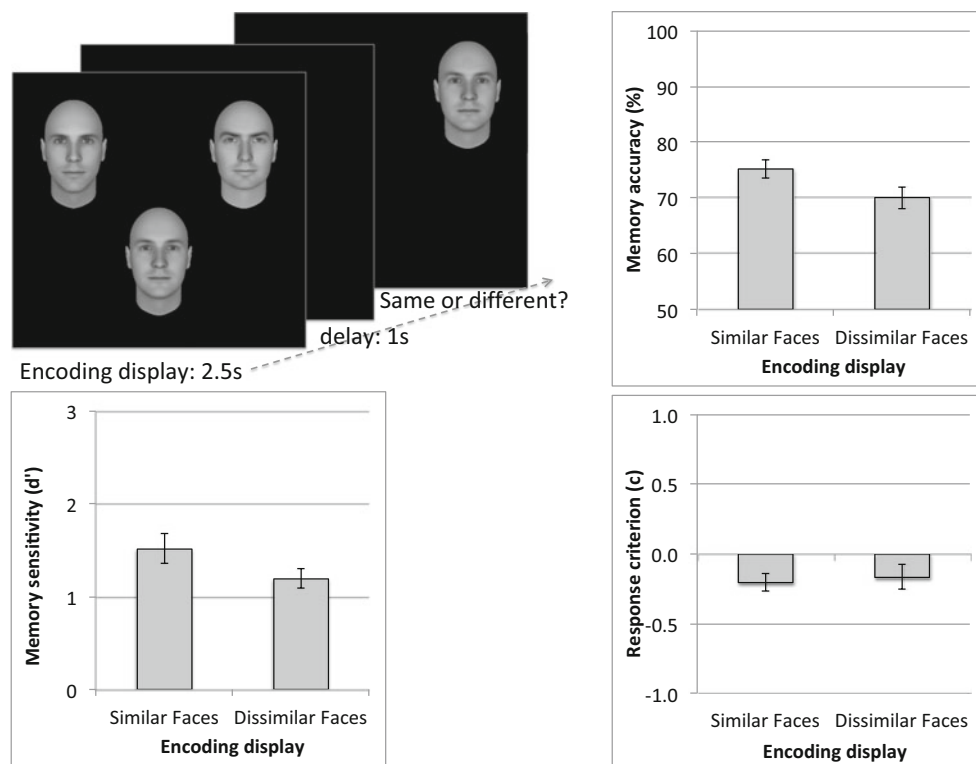

Fig. 2 Trial sequence and results from Experiment 1. *Top left*: trial sequence; *top right*: memory accuracy; *bottom left*: d'; *bottom right*: response criterion. *Error bars* show the between-subject ±1 *SE* of the mean

chose 40 % as the size of change because the difference can be readily perceived (Jiang et al., 2008).

Procedure Participants clicked the display center to initiate each trial. 200 ms later the memory display of three faces was presented for 2.5 s (this duration was adequate for face encoding; Eng, Chen, & Jiang, 2005). The faces occupied three different locations. In Experiment 1, all faces were presented concurrently for 2.5 s. In Experiment 2, the faces were presented sequentially at 700 ms/face, with a 200-ms blank interval between faces. The total duration from the onset of the first face to the offset of the last one was 2.5 s. The locations of the first, second, and third faces were randomly chosen.

After a blank interval of 1 s, the probe display containing a single face appeared. The probe face was in the same location as a randomly selected memory face and remained in view until response. Participants pressed "s" or "d" to judge whether the probe face was the same as the memory face in its location. Tones provided accuracy feedback. We did not emphasize speed.

Design After eight practice trials, participants completed 128 trials in Experiment 1 divided randomly and evenly into two types of memory display (similar or dissimilar) and two probe types (same or different). Because visual working memory is influenced by serial position (Kumar & Jiang, 2005), in Experiment 2 we probed each encoding position equally often. This necessitated the trial number to be multiples of 3. Participants completed 144 trials, divided randomly and evenly into two types of memory display (similar or dissimilar), two probe types (same or different), and three encoding positions.

Results

Experiment 1 (Fig. 2)

Accuracy was significantly higher when the encoding display contained similar rather than dissimilar faces, $t(19) = 2.67$, $p < .05$, Cohen's $d = 1.22$. To test whether similarity influenced memory sensitivity or response criterion, we calculated the sensitivity measure d' and response criterion c (Macmillan & Creelman, 2005). d' is a z score representing the difference between the signal ("target") and noise ("lure") distributions. Response criterion c changes signs depending on whether participants are biased toward reporting "same" (negative values) or "different" (positive values). Results showed significantly higher d' for similar faces than dissimilar faces, $t(19) = 2.10$, $p < .05$, Cohen's $d = 0.96$, with no difference in response criterion for similar and dissimilar conditions, $t(19) = 0.41$, $p > .50$.

Experiment 2 (Fig. 3)

Experiment 1 found a similarity advantage in accuracy, without a change in response criterion. One difference between its procedure and Kahana and Sekuler's previous work was that we presented stimuli simultaneously. Experiment 2 presented stimuli sequentially to examine whether similarity would then have an effect on response criterion. Sequential presentation also allowed us to examine the locus of the similarity advantage. Multiple, simultaneously presented stimuli are known to compete for neuronal representation (Desimone & Duncan, 1995). Such competition could conceivably be greater for dissimilar than similar faces during perceptual encoding (Shim, Jiang, & Kanwisher, 2013). If the similarity advantage shown in Experiment 1 was due solely to reduced neural competition for similar faces during perceptual encoding, then it should be largely eliminated in Experiment 2. Results showed that accuracy was significantly higher in the similar than the dissimilar condition, $F(1, 19) = 6.63$, $p < .05$, $\eta_p^2 = .26$, suggesting that the similarity effect had a memory, rather than just an encoding, component. Memory was better for later serial positions, $F(2, 38) = 46.00$, $p < .001$, $\eta_p^2 = .71$ (the linear trend of serial position was significant, $F(1, 19) = 72.78$, $p < .001$, $\eta_p^2 = .79$). Serial position did not interact with similarity, $F < 1$.

Memory sensitivity d' was marginally higher for similar than dissimilar faces, $F(1, 19) = 3.23$, $p = .088$, $\eta_p^2 = .15$, and higher for later serial positions, $F(2, 38) = 29.43$, $p < .001$, $\eta_p^2 = .61$, with no interaction between similarity and serial position, $F < 1$. The weaker statistical significance in d' can be attributed to extreme values: as a z score, d' is sensitive to extreme values in hits or false alarms (such as 100 % or 0 %), which were likely to happen because each temporal position had just 12 trials. Extreme values were reduced when hits and false alarms were calculated across all temporal positions. The resultant d' was significantly higher for similar faces (mean 1.54) than dissimilar faces (mean 1.31), $t(19) = 2.46$, $p < .03$, Cohen's $d = 1.13$. Even with sequential presentation, similarity had no effects on response criterion, $F < 1$.

Combining data from Experiments 1 and 2 revealed significantly higher accuracy and d' for the similar than dissimilar conditions, $p\text{s} < .01$, but no effects of similarity on response criterion, $p > .20$. These effects did not interact with experiment, $F < 1$.

Discussion

Two experiments showed that visual working memory was better for similar faces than dissimilar faces in both simultaneous and sequential presentations. Because sequential
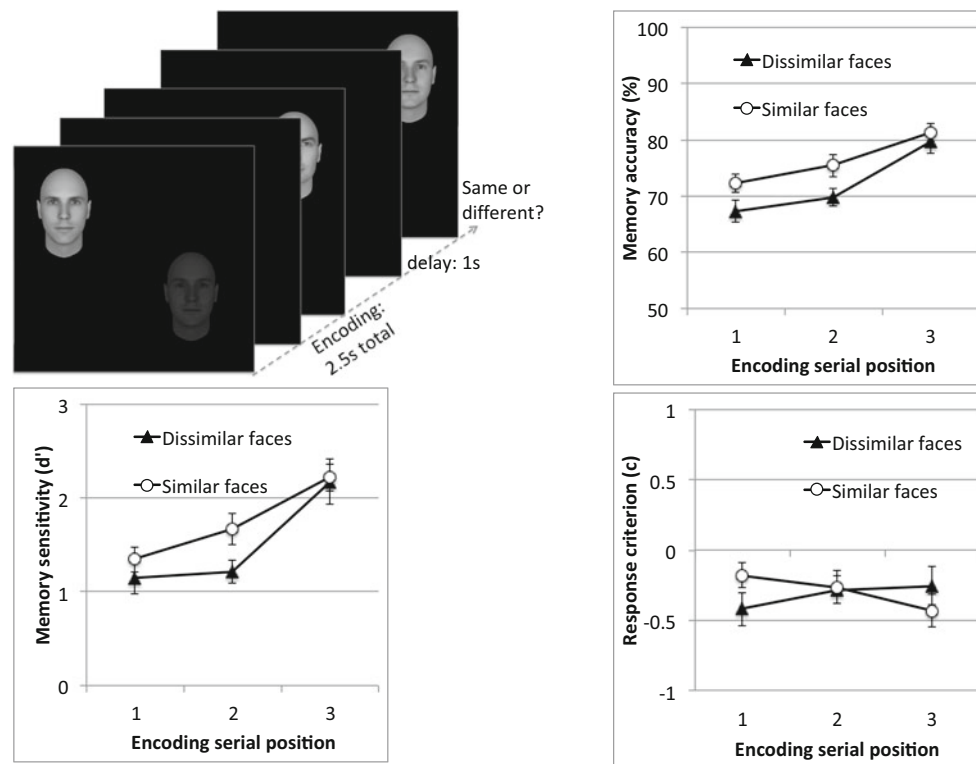
Fig. 3 Trial sequence and results from Experiment 2. *Top left*: trial sequence; *top right*: memory accuracy; *bottom left*: d'; *bottom right*: response criterion. *Error bars* show between-subject ±1 *SE* of the mean

presentation minimized neural competition at the perceptual level, the similarity effect was unlikely due to differential competition for neural representation during encoding. The similarity advantage was consistent with that shown for colors (Lin & Luck, 2009) and line orientations and lengths (Sims et al., 2012). Thus, high degrees of similarity facilitated memory, even when the stimuli were complex. These results constrain the cortical resource theory as a general account of similarity effects (Cohen et al., 2014). They also present an exception to the finding that inter-item similarities affect response criterion (Kahana et al., 2007).

Our data and previous findings suggest two opposing effects of similarity on memory. First, when stimuli are drawn from a confined, contiguous region of similarity space, participants can extract relative properties along a similarity dimension. The reduced featural variance allows participants to have more precise representations of individual memory items (Lin & Luck, 2009; Sims et al., 2012; Swan & Wyble, 2014). According to Sims et al. (2012), reduced feature variance allows processing resources to be concentrated on a smaller region of feature space, increasing precision. In contrast, when stimuli are highly divergent (e.g., different scenes or different faces) the perceptual system cannot extract common perceptual values along a single dimension. Averaging of multiple different faces has to occur in a high-dimensional space (Burton, Jenkins, Hancock, & White, 2005). Though people can extract an "average" face, this does not facilitate memory

for individual faces (de Fockert & Wolfenstein, 2009; Haberman & Whitney, 2007). Mixed-category stimuli yield better memory than single-category stimuli because they rely on widely separated regions in cortical space (Cohen et al., 2014). Thus, reduced featural variance and increased cortical resources are two opposing effects that, under different conditions and for different stimuli, can combine to yield either a similarity advantage or disadvantage.

Unlike Sekuler and Kahana's work, our findings showed that inter-item similarity enhanced memory sensitivity without changing response criterion. Two differences between our study and Kahana and Sekuler's work may explain this discrepancy. First, we presented items at different locations, with the probe appearing at the location of one of the memory items. Participants could use location to selectively retrieve and compare the relevant memory with the probe. In contrast, Kahana and Sekuler presented all memory items, as well as the probe, in the same location. Without location as a retrieval cue participants had to retrieve and compare all memory items with the probe. The two mechanisms important for solving Kahana and Sekuler's task – computing global similarity and extracting inter-item homogeneity – may have less impact in our task. Second, Kahana and Sekuler typically used a large number of stimuli that varied slightly in similarity. Their stimulus condition was analogous to variations in similarity *within* our similar condition. For example, face sets {A40, A50, A60} and {A40, A70, and A100} both belonged to the similar

condition, but the first set was more homogenous than the second.

To examine the two possibilities raised above, we analyzed effects of inter-item similarity within the similar face condition. We used the average distance between face morphs as a measure of mean homogeneity. For example, the mean pairwise difference between faces A30, A40, and A50 is 13.33 %. The mean homogeneity index, calculated on a trial-by-trial basis, fell into six levels ranging from 13.3 % to 46.7 %. Owing to relatively small number of trials per participant, we pooled data from both experiments. Figure 4 displays accuracy as a function of whether the test probe matched or mismatched the memory item.

If participants were less likely to say "same" as homogeneity increased, then greater homogeneity should decrease hit rates while increasing correct rejection rates. An ANOVA on type of response (hits or correct rejections) and homogeneity revealed just the main effect of response type (higher hits than correct rejections, $p < .001$), but no effect of homogeneity, $F(5, 170) = 1.60$, $p > .10$, and no interaction between homogeneity and response type, $F(5, 170) = 1.66$, $p > .10$. These data suggested that the availability of location as a retrieval cue reduced the applicability of Kahana and Sekuler's theory.

Furthermore, the availability of location for retrieval altered the extraction of global similarity. If participants had computed global similarity between the probe and all memory items, then the two unprobed memory items should bias responses. To evaluate this possibility, we calculated the similarity between the unprobed memory items and the probe. We separated trials into two types and measured the proportion of trials in which participants responded "same." In one type the unprobed memory items were highly similar to the probe, with a mean morph difference of 12.5 % (range 5 %–20 %). In another type, the unprobed memory items were less similar to the probe, with a mean morph difference of 32.5 % (range 25 %–40 %). We found that participants were no more likely
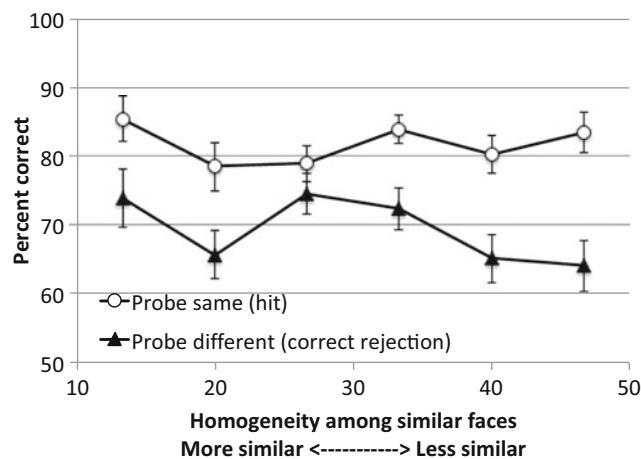
to respond "same" in the first type (51 % "same" response) than in the second (58 %), $p > .95$. This finding suggests that the availability of location as a retrieval cue reduced the impact of ensemble coding.

Conclusions

By testing face morphs, our study expanded the stimulus conditions under which a similarity advantage was observed in visual working memory. We suggest that inter-item similarity has multiple, sometimes opposing, effects on memory. High similarity along a feature dimension may facilitate performance by reducing the noise in memory representation. The advantage for similar stimuli may reverse to a disadvantage when the similar stimuli were themselves highly divergent, such as when comparing single-category with mixed-category stimuli. Future research should consider the multitude of mechanisms that contribute to similarity effects and examine how these mechanisms interact with retrieval.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131. doi:10.1016/j.tics.2011.01.003

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*(4), 362–365. doi:10.1080/14640746608400055

Brady, T. F., & Alvarez, G. A. (2014). No Evidence for a fixed object limit in working memory: Spatial ensemble representations inflate estimates of working memory capacity for complex objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000075

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Burton, A. M., Jenkins, R., Hancock, P. J. B., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology, 51*(3), 256–284. doi:10.1016/j.cogpsych.2005.06.003

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, 111*(24), 8955–8960. doi:10.1073/pnas.1317860111

De Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology, 62*(9), 1716–1722. doi:10.1080/17470210902811249

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222. doi:10.1146/annurev.ne.18.030195.001205

Fig. 4 Memory accuracy as a function of homogeneity among similar faces. *Error bars* show between-subject ±1 *SE* of the mean

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review, 12*(6), 1127–1133.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology: CB, 17*(17), R751–R753. doi:10.1016/j.cub.2007.06.039

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*(3), 683–702.

Jiang, Y. V., Shim, W. M., & Makovski, T. (2008). Visual working memory for line orientations and face identities. *Perception & Psychophysics, 70*(8), 1581–1591. doi:10.3758/PP.70.8.1581

Kahana, M. J., Zhou, F., Geller, A. S., & Sekuler, R. (2007). Lure similarity affects visual episodic recognition: Detailed tests of a noisy exemplar model. *Memory & Cognition, 35*(6), 1222–1232.

Kumar, A., & Jiang, Y. (2005). Visual short-term memory for sequential arrays. *Memory & Cognition, 33*(3), 488–498.

Lin, P.-H., & Luck, S. J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition, 17*(3), 356–372. doi:10.1080/13506280701766313

Luck, S. J., & Hollingworth, A. R. (2008). *Visual memory.* New York, NY: Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: a user's guide. Mahwah, N.J.: Lawrence Erlbaum Associates.

Nosofsky, R. M., & Kantner, J. (2006). Exemplar similarity, study list homogeneity, and short-term perceptual recognition. *Memory & Cognition, 34*(1), 112–124.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & Psychophysics, 64*(7), 1055–1067.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8*(1), 107–113. doi:10.1038/nn1370

Shim, W. M., Jiang, Y. V., & Kanwisher, N. (2013). Redundancy gains in retinotopic cortex. *Journal of Neurophysiology, 110*(9), 2227–2235. doi:10.1152/jn.00175.2013

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review, 119*(4), 807–830. doi:10.1037/a0029856

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics, 76*(7), 2071–2079. doi:10.3758/s13414-014-0690-7

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics, 76*(7), 2136–2157. doi:10.3758/s13414-014-0633-3

Viswanathan, S., Perl, D. R., Visscher, K. M., Kahana, M. J., & Sekuler, R. (2010). Homogeneity computation: How interitem similarity in visual short-term memory alters recognition. *Psychonomic Bulletin & Review, 17*(1), 59–65. doi:10.3758/PBR.17.1.59

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*(1), 48–64.

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review, 10*(1), 80–87.